EDITORS

Hyo-Jeong SO, Ewha Womans University, South Korea
Ma. Mercedes RODRIGO, Ateneo de Manila University, Philippines
Jon MASON, Charles Darwin University, Australia
Antonija MITROVIC, University of Canterbury, New Zealand
ASSOCIATE EDITORS
(in alphabetical order)

Daniel BODEMER, University of Duisburg-Essen, Germany
Weichao CHEN, University of Virginia, USA
Zhi-Hong CHEN, National Taiwan Normal University, Taiwan
Brendan FLANAGAN, Kyoto University, Japan
Marc JANSEN, Hochschule Ruhr West, Germany
Roger NKAMBOU, Université du Québec à Montréal, Canada
Longkai WU, Nanyang Technological University, Singapore
MESSAGE FROM THE CONFERENCE CHAIR

On behalf of the organizing committee, I would like to welcome all participants of the 28th International Conference on Computers in Education (ICCE) 2020, the flagship conference series of the Asia-Pacific Society for Computers in Education (APSCE). As a consequence of COVID-19, this is the first virtual ICCE conference. We originally planned to meet in Darwin, Australia, and enjoy the hospitality of Jon Mason and his team. Alas, that was not to be. We thank Jon Mason for all his efforts on organizing the conference in Australia, and also the time he has put later on into making this virtual conference possible.

We will all remember 2020 for a long time. The year has brought a lot of anxiety, uncertainty and changes. I would like to thank our standing committee, for being flexible and finding solutions to challenging problems we faced. Our appreciation goes to Pham-Duc Tho, the Managing Secretary of APSCE, who graciously accepted the tasks that are usually taken by the local organizing team. Tho and his team took care of the ICCE 2020 Web site, all communication with participants, registrations, and the hands-on organization of sessions. The task was huge, and we are all indebted to Pham-Duc.

My sincere appreciation goes to Hyo-Jeong So, Didith Rodrigo and Jon Mason, the chair and co-chairs of the International Program Committee respectively. They have put an enormous amount of time in making sure that we have an excellent programme at ICCE 2020. My gratitude goes to the chairs of the seven sub-conferences, organizers of workshops, tutorials, panels, WIPP, DSC, ES, posters, ECW. And of course, our sincere thanks to all authors, reviewers, presenters, Doctoral students, and other participants. I would also like to thank our consultants, Lung-Hsiang Wong, Maiga Chang, Fu-Yun Yu and Juling Shih, for sharing their wisdom and advising us along the way.

Four outstanding keynote speakers will share their insights across varying areas in the field of computers in education. They are (1) Peter Goodyear from University of Sydney, Australia, who will present his reflections on learning, technology and design; (2) Vania Dimitrova from the University of Leeds, UK, who will present her ideas on intelligent mentoring systems; (3) Sasha Barab from the Arizona State University, USA, who will present his views on learning to thrive; and (4) Lung-Hsiang Wong from the Nanyang Technological University, Singapore, who will talk about cross-fertilization between different research orientations with the focus on mobile learning.

Antonija MITROVIC
Conference Chair
University of Canterbury
New Zealand
There will also be three equally inspiring theme-based invited speeches. Emma Mercier from the University of Illinois at Urbana Champaign, USA, will talk about supporting collaborative learning in classrooms. Chengjiu Yin from Kobe University, Japan, will talk about learning analytics applied to e-book reading logs. Ting-Chia Hsu, from the National Taiwan Normal University, will talk about studies on learning language from mobile applications.

I hope the participants will find the conference invigorating, relevant and enjoyable, and that will be able to meet face-to-face in 2021!
MESSAGE FROM THE INTERNATIONAL PROGRAM COORDINATION CHAIRS

The International Conference on Computers in Education (ICCE) is an annual conference series encompassing diverse issues related to the use of Information and Communication Technology (ICT) in various settings of education, organized by the Asia-Pacific Society for Computers in Education (APSCE).

Due to the COVID-19 pandemic, ICCE 2020 is held virtually from November 23 to November 27, 2020. While it is not possible to gather together in a physical conference place, we believe that the virtual space offers a flexible platform where researchers around the world share their latest research findings, insights and work-in-progress ideas that can advance the field of computers in education.

While ICCE 2020 is shifting to a fully online mode, the tradition of the previous ICCEs was followed as the meta-conference with seven sub-conference programs specializing specific themes:

- C1: ICCE Sub-Conference on Artificial Intelligence in Education/Intelligent Tutoring System (AIED/ITS)
- C2: ICCE Sub-Conference on Computer-supported Collaborative Learning (CSCL) and Learning Sciences (LS)
- C3: ICCE Sub-Conference on Advanced Learning Technologies (ALT), Learning Analytics, Platforms and Infrastructure
- C4: ICCE Sub-Conference on Classroom, Ubiquitous, and Mobile Technologies Enhanced Learning (CUMTEL)
- C5: ICCE Sub-Conference on Educational Gamification and Game-based Learning (EGG)
- C6: ICCE Sub-Conference on Technology Enhanced Language Learning (TELL)
- C7: ICCE Sub-Conference on Practice-driven Research, Teacher Professional Development and Policy of ICT in Education (PTP)

The International Program Committee is led by a strong and dedicated team, which includes the Conference Chair, the Program Coordination Chair and Co-Chair, seven executive Sub-Conference Chairs and 255 experts in the field of Computers in Education from 39 different countries or economies. Former ICCE local organizing and program coordination chairs have played important roles as consultants in overseeing the organization process of this conference.
ICCE 2020 received a total of 159 submissions\(^1\) (115 full, 36 short, and 8 posters) from 27 different countries or economies. Top three countries with the highest number of submissions are Japan, Australia and China. Submissions were also received from the Middle East, Europe, America and Africa, which signals the international interest toward ICCE 2020. Table 1 provides the submissions statistics by the country where the first author comes from:

### Table 1. Submission statistics by country (based on first author’s country)

<table>
<thead>
<tr>
<th>Countries or Economies</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>14</td>
</tr>
<tr>
<td>Brazil</td>
<td>1</td>
</tr>
<tr>
<td>Canada</td>
<td>2</td>
</tr>
<tr>
<td>China</td>
<td>14</td>
</tr>
<tr>
<td>Finland</td>
<td>1</td>
</tr>
<tr>
<td>Germany</td>
<td>2</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>13</td>
</tr>
<tr>
<td>India</td>
<td>3</td>
</tr>
<tr>
<td>Indonesia</td>
<td>4</td>
</tr>
<tr>
<td>Iran</td>
<td>1</td>
</tr>
<tr>
<td>Japan</td>
<td>40</td>
</tr>
<tr>
<td>Malaysia</td>
<td>4</td>
</tr>
<tr>
<td>New Zealand</td>
<td>6</td>
</tr>
<tr>
<td>Philippines</td>
<td>12</td>
</tr>
<tr>
<td>Poland</td>
<td>4</td>
</tr>
<tr>
<td>Singapore</td>
<td>7</td>
</tr>
<tr>
<td>Spain</td>
<td>2</td>
</tr>
<tr>
<td>Sweden</td>
<td>2</td>
</tr>
<tr>
<td>Taiwan</td>
<td>12</td>
</tr>
<tr>
<td>Tunisia</td>
<td>2</td>
</tr>
<tr>
<td>Turkey</td>
<td>1</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>1</td>
</tr>
<tr>
<td>United States</td>
<td>5</td>
</tr>
</tbody>
</table>

All papers were subjected to a rigorous review process by at least three reviewers from the respective Sub-Conference program committees. After the reviews were completed, a meta-review was provided for each paper. In total, 499 reviews and meta-reviews were received. After the discussion period within the individual program committees led by the Sub-Conference Executive Chairs and Co-Chairs, recommendations were made to the Program Coordination Committee Chair and Co-Chair, who oversaw the review process and quality for all

---

\(^1\) Six papers were withdrawn after the review process was complete.
Sub-Conferences. This resulted in 28 full papers, 60 short papers, and 22 posters accepted across seven Sub-Conferences. The overall acceptance rate for full papers is 25.70%, which reflects our efforts to continue the maintenance of the quality of presentations at ICCE 2020. The complete statistics of paper acceptance is shown in Table 2.

Table 2. Paper Acceptance Statistics

<table>
<thead>
<tr>
<th></th>
<th>Submission</th>
<th>Submit as Full</th>
<th>Accepted as Full</th>
<th>Full %</th>
<th>Accepted as Short</th>
<th>Accepted as Poster</th>
<th>Overall %</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1 - AIED/ITS</td>
<td>22</td>
<td>20</td>
<td>4</td>
<td>20.00</td>
<td>7</td>
<td>4</td>
<td>68.18</td>
</tr>
<tr>
<td>C2 - CSCL/LS</td>
<td>21</td>
<td>16</td>
<td>5</td>
<td>31.25</td>
<td>6</td>
<td>4</td>
<td>71.43</td>
</tr>
<tr>
<td>C3 - ALT/LA</td>
<td>32</td>
<td>22</td>
<td>6</td>
<td>27.27</td>
<td>16</td>
<td>1</td>
<td>71.88</td>
</tr>
<tr>
<td>C4 - CUMTEL</td>
<td>13</td>
<td>7</td>
<td>2</td>
<td>28.57</td>
<td>5</td>
<td>2</td>
<td>69.23</td>
</tr>
<tr>
<td>C5 - EGG</td>
<td>14</td>
<td>12</td>
<td>3</td>
<td>25.00</td>
<td>6</td>
<td>1</td>
<td>71.43</td>
</tr>
<tr>
<td>C6 - TELL</td>
<td>24</td>
<td>13</td>
<td>3</td>
<td>23.08</td>
<td>10</td>
<td>5</td>
<td>75.00</td>
</tr>
<tr>
<td>C7 - PTP</td>
<td>27</td>
<td>19</td>
<td>5</td>
<td>26.32</td>
<td>10</td>
<td>5</td>
<td>74.07</td>
</tr>
<tr>
<td>Total</td>
<td>153</td>
<td>109</td>
<td>28</td>
<td>25.70</td>
<td>60</td>
<td>22</td>
<td>71.90</td>
</tr>
</tbody>
</table>

In addition to the main program with seven sub-conferences, ICCE 2020 includes various program components, such as Keynote Speeches, Theme-based Invited Speeches, Workshops, Tutorials, Work-in-Progress Posters (WIPP), Extended Summary (ES), Doctoral Student Consortia (DSC), and Early Career Workshop (ECW). All the papers in these program components are compiled and published in a separate volume with its own ISBN. Pre-conference events are held on the first two days of the conference, including 13 workshops, one tutorial, DSC, ECW, APSCE Student Wing Workshop, and SIG community building sessions.

In closing, we would like to thank those who have contributed to making ICCE 2020 a successful conference. First of all, we would like to thank all the paper authors for choosing ICCE 2020 as an outlet to present their research. We would also like to thank the IPC Executive Chairs/Co-Chairs and members, who undertook the responsibility of reviewing and selecting papers that represent research of high quality. Specially thanks to our Keynote and Invited Speakers for accepting our invitations and sharing inspiring research with the ICCE 2020 participants.

ICCE 2020 will be remembered as our first-ever virtual conference. We hope that all participants will find the activities in ICCE 2020 interesting and inspiring, and have opportunities to meet old friends and establish new professional collaborations in a virtual space. Thank you for all your support during these challenging times.
ORGANIZATION

Organized by: Asia Pacific Society for Computers in Education

Standing Committee
- Conference Chair
  Antonija MITROVIC, University of Canterbury, New Zealand
- International Program Coordination Chair
  Hyo-Jeong SO, Eawha Womans University, South Korea
- International Program Coordination Co-Chairs
  Ma. Mercedes T. RODRIGO, Ateneo de Manila University, Philippines
  Jon MASON, Charles Darwin University, Australia
- Consultants
  Maiga CHANG, Athabasca University, Canada
  Lung Hsiang WONG, Nanyang Technological University, Singapore
  Fu-Yun YU, National Cheng Kung University, Taiwan
  Juling SHIH, National University of Tainan, Taiwan

Sub Conferences
C1: Sub-Conference on Artificial Intelligence in Education/Intelligent Tutoring Systems (AIED/ITS)
- PC Executive Chair:
  Roger NKAMBOU, Université du Québec à Montréal, Canada
- PC Executive Co-Chairs:
  Philippe FOURNIER-VIGER, Harbin Institute of Technology (Shenzhen), China
  John STAMPER, Carnegie Mellon University, USA
  Kazuhisa SETA, Osaka Prefecture University, Japan

C2: Sub-Conference on Computer-supported Collaborative Learning (CSCL) and Learning Sciences
- PC Executive Chair:
  Daniel BODEMER, University of Duisburg-Essen, Germany
- PC Executive Co-Chairs:
  Elizabeth KOH, Nanyang Technological University, Singapore
  Sahana MURTHY, Indian Institute of Technology Bombay, India
  Kate THOMPSON, Queensland University of Technology, Australia
  Camillia MATUK, New York University, USA
C3: Sub-Conference on Advanced Learning Technologies, Learning Analytics, Platforms and Infrastructure

- **PC Executive Chair:**
  Brendan FLANAGAN, Kyoto University, Japan

- **PC Executive Co-Chairs:**
  Ramkumar RAJENDRAN, Indian Institute of Technology Bombay, India
  Gökhan AKÇAPINAR, Hacettepe University, Turkey
  Solomon OYELERE, University of Eastern Finland, Finland

C4: Sub-Conference on Classroom, Ubiquitous and Mobile Technologies Enhanced Learning (CUMTEL)

- **PC Executive Chair:**
  Longkai WU, Nanyang Technological University, Singapore

- **PC Executive Co-Chairs:**
  Kaushal Kumar BHAGAT, Indian Institute of Technology, Kharagpur, India
  Jingyun WANG, Kyushu University, Japan
  Jun LIU, Capital Normal University, China
  Andrea VALENTE, University of Southern Denmark, Denmark

C5: Sub-Conference on Educational Gamification and Game-based Learning (EGG)

- **PC Executive Chair:**
  Zhi-Hong CHEN, National Taiwan Normal University, Taiwan

- **PC Executive Co-Chairs:**
  Hercy N. H. CHENG, Central China Normal University, China
  Yi-Hsuan WANG, Tamkang University, Taiwan
  Rhodora ABADIA, University of South Australia, Australia
  Yi-Chun HONG, Arizona State University, USA

C6: Sub-Conference on Technology Enhanced Language Learning (TELL)

- **PC Executive Chair:**
  Weichao CHEN, University of Virginia, USA

- **PC Executive Co-Chairs:**
  Michael THOMAS, Liverpool John Moores University, UK
  Yun WEN, Nanyang Technological University, Singapore
  Agnieszka PALALAS, Athabasca University, Canada
  Lynde TAN, Western Sydney University, Australia

C7: Sub-Conference on Practice-driven Research, Teacher Professional Development and Policy of ICT in Education (PTP)

- **PC Executive Chair:**
  Marc JANSEN, Hochschule Ruhr West, Germany

- **PC Executive Co-Chairs:**
  Ying ZHAN, The Education University of Hong Kong, Hong Kong
Dan KOHEN-VACS, Holon Institute of Technology, Israel
Shitanshu MISHRA, Vanderbilt University, USA

Workshop & Interactive Events
- Chair:
  Patcharin PANJABUREE, Mahidol University, Thailand
- Co-Chairs:
  Swapna GOTTPATI, Singapore Management University, Singapore
  Charoenchai WONGWATKIT, Mae Fah Luang University, Thailand

Tutorials
- Chair:
  Ahmed TLILI, Smart Learning Institute of Beijing Normal University, China
- Co-Chairs:
  Rustam SHADIEV, Nanjing Normal University, China

Work-In-Progress Posters (WIPP)
- Chair:
  Michelle P. BANAWAN, Arizona State University, USA
- Co-Chairs:
  Jenilyn L. AGAPITO, Ateneo de Manila University, Philippines
  Kazuaki KOJIMA, Teikyo University, Japan

Doctoral Student Consortium (DSC)
- Chair:
  Tatsunori MATSUI, Waseda University, Japan
- Co-Chairs:
  Morris JONG, The Chinese University Hong Kong, Hong Kong
  Bo JIANG, East China Normal University, China

Early Career Workshop (ECW)
- Chair:
  Ma. Mercedes T. RODRIGO, Ateneo de Manila University, Philippines
- Co-Chairs:
  Hiroaki OGATA, Kyoto University, Japan
  Mas Nida BT MD KHAMBARI, Universiti Putra Malaysia, Malaysia
  Thepchai SUPNITHI, National Electronics and Computer Technology Center, Thailand

Panels
- Chair:
  Ali DEWAN, Athabasca University, Canada
- Co-Chairs:
  Han-Yu SUNG, National Taipei University of Nursing and Health Science, Taiwan
Extended Summaries (ES)

- **Chair:**
  Chengjiu YIN, Kobe University, Japan
- **Co-Chairs:**
  Ping Li, The Hong Kong Polytechnic University, Hong Kong

Merit Scholarships

- **Chair:**
  Madathil Warriem JAYAKRISHNAN, Indian Institute of Technology Madras, India
- **Co-Chairs:**
  Mohammed Nehal HASNINE, Tokyo University of Agriculture and Technology, Japan

Special Interest Groups (SIGs)

- **S1:** Artificial Intelligence in Education/Intelligent Tutoring Systems/Adaptive Learning (AIED/ITS/AL)
  Michelle P. BANAWAN, Arizona State University, USA
- **S2:** Computer-supported Collaborative Learning (CSCL) and Learning Sciences
  Chew Lee TEO, Nanyang Technological University, Singapore
- **S3:** Advanced Learning Technologies (ALT), Open Contents, and Standards
  Jin Gon SON, Korea National Open University, Korea
- **S4:** Classroom, Ubiquitous and Mobile Technologies Enhanced Learning (CUMTEL)
  Ting-Chia HSU, National Taiwan Normal University, Taiwan
- **S5:** Educational Gamification and Game-based Learning (EGG)
  Rita KUO, New Mexico Institute of Mining and Technology, Taiwan
- **S6:** Technology Enhanced Language Learning (TELL)
  Yoshiko GODA, Kumamoto University, Japan
- **S7:** Practice-driven Research, Teacher Professional Development, and Policy of ICT in Education (PTP)
  Sahana MURTHY, Indian Institute of Technology Bombay, India
- **S8:** Development of Information and Communication Technology in the Asia-Pacific Neighborhood (DICTAP)
  Bo JIANG, Zhejiang University of Technology, China
- **S9:** Educational Use of Problems/Questions in Technology-Enhanced Learning
  Kazuaki KOJIMA, Teikyo University, Japan
- **S10:** Learning Analytics and Educational Data Mining
  Brendan FLANAGAN, Kyoto University, Japan
- **S11:** Computational Thinking Education & STEM Education (CTE & STEM)
  Siu Cheung KONG, The Education University of Hong Kong, Hong Kong

C1 PC Members

- Sagaya Amalathas, Taylors University, Malaysia
- Ryan Baker, University of Pennsylvania, USA
● Michelle Banawan, Ateneo de Davao University, Philippines
● Kritya Bunchongchit, Mahidol University International College, Thailand
● Olga C. Santos, aDeNu Research Group, UNED, Spain
● Chih-Yueh Chou, Yuan Ze University, Taiwan
● Benedict Du Boulay, University of Sussex, UK
● Claude Frasson, University of Montreal, Canada
● Yusuke Hayashi, Hiroshima University, Japan
● Bastiaan Heeren, Open University, Netherlands
● Tomoya Horiuchi, Kobe University, Japan
● Sharon Hsiao, Arizona State University, USA
● Akihiro Kashihara, The University of Electro-Communications, Japan
● Tomoko Kojiri, Kansai University, Japan
● Tatsuhiko Konishi, Shizuoka University, Japan
● Nguyen-Thanh Le, Humboldt-Universität zu Berlin, Germany
● Noboru Matsuda, North Carolina State University, USA
● Tatsunori Matsui, Waseda University, Japan
● Antonija Mitrovic, University of Canterbury, New Zealand
● Kazuhisa Miwa, Nagoya University, Japan
● Riichiro Mizoguchi, Japan Advanced Institute of Science and Technology, Japan
● Helen Pain, The University of Edinburgh, UK
● Jose Luis Perez De La Cruz, Universidad de Malaga, Spain
● Elvira Popescu, University of Craiova, Romania
● Ma. Mercedes T. Rodrigo, Ateneo de Manila University, Philippines
● Thepchai Supnithi, National Electronics and Computer Technology Center, Thailand
● Ange Tato, Université du Québec à Montréal, Canada

C2 PC Members

● Peter Reimann, The University of Sydney, Australia
● Chew Lee Teo, Nanyang Technological University, Singapore
● Sakinah Alhadad, Griffith University, Australia
● Jürgen Buder, Leibniz-Institut für Wissensmedien, Germany
● Sanjay Chandrasekharan, Homi Bhabha Centre for Science Education, TIFR, India
● Andrew Clayphan, The University of Sydney, Australia
● Chandan Dasgupta, Indian Institute of Technology Bombay, India
● Pryce Davis, University of Nottingham, UK
● Amanda Dickes, Gulf of Maine Research Institute, USA
● Julia Eberle, Ruhr-University Bochum, Germany
● Heisawn Jeong, Hallym University, Korea
● Beaumie Kim, University of Calgary, Canada
● Simon Knight, University of Technology, Sydney, Australia
● Rose Liang, National University of Singapore, Singapore
● Pei-Yi Lin, National Kaohsiung Normal University, Taiwan
● Chen-Chung Liu, National Central University, Taiwan
● Chee Kit Looi, Nanyang Technological University, Singapore
● Sven Manske, University of Duisburg-Essen, Germany
● Roberto Martínez-Maldonado, Monash University, Australia
● Johanna Pöysä-Tarhonen, University of Jyväskylä, Finland
● Nikol Rummel, Ruhr-University Bochum, Germany
● Lenka Schnaubert, University of Duisburg-Essen, Germany
● Kristen Searle, Utah State University, USA
● Hillary Swanson, Northwestern University, USA
● Seng Chee Tan, Nanyang Technological University, Singapore
● Hongzhi Yang, The University of Sydney, Australia

C3 PC Members
● Khalid Khan, Charles Darwin University, Australia
● Mei-Rong Alice Chen, National Taiwan University of Science and Technology, Taiwan
● Tomczyk Łukasz, Pedagogical University of Cracow, Poland
● Teo Chewlee, National Institute of Education, Singapore
● Barros Gabriel, Universidad del Azuay, Ecuador
● Mohammed Saqr, University of Eastern Finland, Finland
● Kwok Lam For, City University of Hong Kong, Hong Kong
● Silva César Villacís, Universidad de las Fuerzas Armadas, Ecuador
● Eliseo Maria Amelia, Universidade Presbiteriana Mackenzie, Brazil
● Mitra Ritayan, Indian Institute of Technology Bombay, India
● Silveira Ismar Frango, Universidade Presbiteriana Mackenzie, Brazil
● Sun Jerry Chih-Yuan, National Chiao Tung University, Taiwan
● Azcarraga Judith, De La Salle University, Philippines
● Er Erkan, Universidad de Valladolid, Spain
● Mason Jon, Charles Darwin University, Australia
● Kokoç Mehmet, Karadeniz Technical University, Turkey
● Hoel Tore, Oslo Metropolitan University, Norway
● Wang Minhong, The University of Hong Kong, Hong Kong
● Miao Yongwu, University Duisburg-Essen, Germany
● Vuorikari Riina, Joint Research Centre European Commission, Spain
● Martins Valéria Farinazzo, Universidade Presbiteriana Mackenzie, Brazil
● Kolog Emmanuel Awuni, University of Ghana Business School, Ghana
● Balogun Oluwafemi Samson, University of Eastern Finland, Finland
● Caieiro Rodríguez Manuel, University of Vigo, Spain
● Stanger Nigel, University of Otago, New Zealand
● Motz Regina, Universidad de la República, Uruguay
● Olsen Jennifer, École Polytechnique Fédérale de Lausanne, Switzerland
● Cibelle Amato, Universidade Presbiteriana Mackenzie, Brazil
● Chen Yang-Hsueh, National Chengchi University, Taiwan
● Kohen-Vacs Dan, Holon Institute of Technology, Israel
● Gottipati Swapna, Singapore Management University, Singapore
• Anido Rifon Luis, Universidade de Vigo, Spain
• Amado-Salvatier Héctor R., Universidad Galileo, Guatemala
• Mohammed Nehal Hasnine, Hosei University, Japan

C4 PC Members
• Chiu-pin Lin, National Tsing Hua University, Taiwan
• Yanjie Song, The Education University of Hong Kong, Hong Kong
• Yun Wen, Nanyang Technological University, Singapore
• Chih-Ming Chu, National ILan University, Taiwan
• Ting-Ting Wu, National Yunlin University of Science and Technology, Taiwan
• Noriko Uosaki, Osaka University, Japan
• Yih-Ruey Juang, Jinwen University of Science and Technology, Taiwan
• Jun-Ming Su, National University of Tainan, Taiwan
• Martina Holenko Dlab, University of Rijeka, Croatia
• Ivana Bosnić, University of Zagreb, Croatia
• M. Carmen Juan Lizandra, Universitat Politècnica de València, Spain
• Igor Mekterović, University of Zagreb, Croatia
• Tai-Chien Kao, National Dong Hwa University, Taiwan
• I-Wen Huang, National University of Tainan, Taiwan
• Gwo-Haur Hwang, National Yulin Technology University, Taiwan
• Chen-Yu Lee, Ling Tung University, Taiwan
• Kuo-Liang Ou, National Tsing Hua University, Taiwan
• Tzu-Chi Yang, National Chiao Tung University, Taiwan
• Ben Chang, National Central University, Taiwan
• Silin Wei, Hangzhou Normal University, China
• Yuen-Hsien Tseng, National Taiwan Normal University, Taiwan
• Chi-Chen Chang, National Taiwan Normal University, Taiwan
• Kai-Hsiang Yang, National Taipei University of Education, Taiwan
• Gwo-Jen Hwang, National Taipei University of Science and Technology, Taiwan
• Haiguang Fang, Capital Normal University, China
• Chiu-Lin Lai, National Taipei University of Education, Taiwan
• Feng-Kuang Chiang, Shanghai Normal University, China
• Xianmin Yang, Jiangsu Normal University, China
• Jing Leng, East China Normal University, China
• Guang Chen, Beijing Normal University, China
• Fati Wu, Beijing Normal University, China
• Jianli Jiao, South China Normal University, China
• Xuefeng Wei, Ludong University, China
• Ping He, Tianjin University, China
• Jing Leng, East China Normal University, China
• Huiying Cai, Jiangnan University, China
• Bian Wu, East China Normal University, China
• Zhihong Wan, The Education University of Hong Kong, Hong Kong
• Mi Song Kim, The University of Western Ontario, Canada
• Luo Ma, East China Normal University, China
• Daner Sun, The Education University of Hong Kong, Hong Kong
• Chengjiu Yin, Kobe University, Japan
• Fuhua (Oscar) Lin, Athabasca University, Canada
• Ivica Boticki, University of Zagreb, Croatia
• Su Cai, Beijing Normal University, China
• Kathryn MacCallum, University of Canterbury, New Zealand

C5 PC Members
• Chung-Yuan Hsu, National Pingtung University of Science and Technology, Taiwan
• Hiroyuki Mitsuhara, Tokushima University, Japan
• Sheng-Kai Yin, Cheng Shiu University, Taiwan
• Anurag Deep, Indian Institute of Technology Bombay, India
• Demetrios Sampson, Curtin University, Australia
• Ahmed Tili, Smart Learning Institute of Beijing Normal University, China
• Liz Bacon, Abertay University, Scotland
• Morris Jong, The Chinese University of Hong Kong, Hongkong
• Chong-Shiuh Koong, National Taichung University, Taiwan
• Kaoru Sumi, Future University Hakodate, Japan
• Ben Chang, National Central University, Taiwan
• Susan Gwee, English Language Institute of Singapore, Singapore
• Wing-Kwong Wong, National Yunlin University of Science & Technology, Taiwan
• Tzu-Chao Chien, National Central University, Taiwan
• Shu-Yuan Tao, Takming University of Science and Technology, Taiwan
• Toshihiro Hayashi, Kagawa University, Japan
• Di Zou, The Education University of Hong Kong, Hong Kong
• Michal Ptaszynski, Kitami Institute of Technology, Japan
• Yi-Lung Lin, China University of Technology, Taiwan
• Chih-Ming Chu, National Ilan University, Taiwan
• Nian-Shing Chen, National Yunlin University of Science and Technology, Taiwan
• Chang-Yen Liao, National Taipei University of Nursing and Health Sciences, Taiwan
• Ming-Puu Chen, National Taiwan Normal University, Taiwan
• Saurabh Mehta, Vidyalankar Institute of Technology, India
• Liang-Yi Li, National Taiwan Normal University, Taiwan
• Ju-Ling Shih, National Central University, Taiwan
• Chiung-Sui Chang, Tamkang University, Taiwan
• Junjie Shang, Peking University, China
• George Ghinea, Brunel University, UK
• Meng-Jung Tsai, National Taiwan Normal University, Taiwan
• Jianhua Wu, Central China Normal University, China
C6 PC Members
- Alex Boulton, University of Lorraine, France
- Mei-Mei Chang, National Pingtung University of Science and Technology, Taiwan
- Ying-Hsueh Cheng, National Pingtung University of Science and Technology, Taiwan
- Jozef Colpaert, University of Antwerp, Belgium
- Yoshiko Goda, Kumamoto University, Japan
- Robert Godwin-Jones, Virginia Commonwealth University, USA
- Yanhui Han, The Open University of China, China
- Chia-Ling Hsieh, National Taiwan Normal University, Taiwan
- Jiyou Jia, Peking University, China
- Ho Cheong Lam, The Education University of Hong Kong, Hong Kong
- Jiahang Li, Michigan State University, USA
- Meei-Ling Liaw, National Taichung University, Taiwan
- Misato Oi, Kyushu University, Japan
- Yuichi Ono, University of Tsukuba, Japan
- Takafumi Utashiro, Hokkai-Gakuen University, Japan
- Jane Vinther, University of Southern Denmark, Denmark
- Limei Zhang, Nanyang Technological University, Singapore
- Shenglan Zhang, Iowa State University, USA
- Chunping Zheng, Beijing University of Posts and Telecommunications, China
- Di Zou, The Education University of Hong Kong, Hong Kong
- Shijuan Liu, Indiana University of Pennsylvania, USA
- Chien-han Chen, Tamkang University of Taiwan, Taiwan
- Wanwisa Wannaipat, Khon Kaen University, Thailand
- Phil Hubbard, Stanford University, USA
- Hui-Chin Yeh, National Yunlin University of Science and Technology, Taiwan
- Zahra Shahsavari, Shiraz University of Medical Sciences, Iran
- Lizeng Huang, University of Virginia, USA
- Sandra Gudino, Tecnológico de Monterrey, Mexico
- Liliana Cuesta Medina, Universidad de La Sabana, Columbia
- Qi Li, Jingle Magic (Beijing) Technology Co., China
- Zhenzhen Chen, Beijing University of Posts and Telecommunications, China
- Chaoran Wang, Indiana University Bloomington, USA
- Xin Chen, Indiana University Bloomington, USA
- Jui-Hsin Renee Hung, Indiana University Bloomington, USA

C7 PC Members
- Lee Shushing, National Institute of Education, Singapore
- Sun Daner, The Education University of Hong Kong, Hong Kong
- Su Luan Wong, Universiti Putra, Malaysia
- Tore Hoel, Oslo Metropolitan University, Norway
- Jayakrishnan M. Warriem, Indian Institute of Technology, India
- Marcelo Milrad, Linnaeus University, Sweden
• Jan Pawlowski, University of Applied Sciences, Germany
• Chee-Kit Looi, National Institute of Education, Singapore
• Gökhan Akçapınar, Hacettepe University, Turkey
• Boticka Ivicki, University of Zagreb, Croatia
• Rwitajit Majumdar, Kyoto University, Japan
• Peter Seow, Nanyang Technological University, Singapore
• Tessy Cerratto-Pargman, Stockholm University, Sweden
• Meital Amzalag, Holon Institute of Technology, Israel
• Tamar Ronen Fuhrmann, Columbia University, New-York
• Tamar Inbal Shamir, The Israeli Open University, Israel
• Aditi Kothiyal, École polytechnique fédérale de Lausanne, Switzerland
• Veenita Shah, Indian Institute of Technology, India
• Yogendra Pal, Indian Institute of Technology, India
• Jennifer Olsen, École polytechnique fédérale de Lausanne, Switzerland
• Ma Luo, Ningbo University, China
• Wai Man Winnie Lam, The Education University of Hong Kong, Hong Kong
APSCE FELLOWS PROGRAM

Founded in 2019, the APSCE Fellowship recognizes outstanding members of the Asia-Pacific Society for Computers in Education (APSCE) in the field of computers in education. The title of APSCE fellow indicates, (1) Sustained and distinguished academic contributions to the advancement of research in the field of computers in education at the international level; (2) A strong track record in academic networking and services within the Asia-Pacific region.

The fellowship is for life, whose names shall be indicated on the APSCE website permanently. Furthermore, the APSCE fellows are entitled to complimentary lifetime voting APSCE memberships.

The number of new fellows named each year shall be capped at five (5). An APSCE Fellow must be an existing APSCE member in the year he or she is inducted.

The inaugural cohort of the APSCE Fellowship consists of the three existing APSCE Honorary Executive Committee (EC) members. Subsequently, the APSCE President, the APSCE Award Subcommittee Chair and the Honorary EC members formed the APSCE Fellow Committee to select additional fellows. After the first year (2019), the existing APSCE Fellows, the APSCE President and the Award Subcommittee Chair shall form the APSCE Fellow Committee each year to select new fellows. The APSCE President and the Award Subcommittee Chair are not eligible for APSCE Fellow inductions in the year in which they are serving as APSCE Fellow Committee members.

The full APSCE Fellowship guidelines is available on https://tinyurl.com/y3xmo7n6

The inaugural cohort of APSCE Fellows are (in alphabetical order):

- Tak-Wai CHAN (Taiwan)
- H. Ulrich HOPPE (Germany)
- Chee-Kit LOOI (Singapore)
- Riichiro MIZOGUCHI (Japan)

The two APSCE Fellows inducted this year (2020) are (in alphabetical order):

- Gautam BISWAS (USA)
- Siu Cheung KONG (Hong Kong)
Tak-Wai CHAN

Chair Professor, Graduate Institute of Networked Learning
National Central University, Taiwan
Website: http://chan.lst.ncu.edu.tw/

Tak-Wai Chan is Chair Professor of the Graduate Institute of Network Learning Technology at National Central University in Taiwan. In the late Eighties, he pioneered research on virtual learning companions. In 1990, his team developed possibly the earliest networked learning system dedicated to support collaborative learning and competitive learning games. He continued to work on various networked learning models, intelligent future classrooms, and digital game-based learning in the Nineties. In 2000, his team built an online learning society called EduCity. By 2003, EduCity had 2,500 schools and 1.5 million online participants before it was transferred to a telecom company. Throughout the 2000’s, he worked on mobile and ubiquitous learning, one-to-one technology enhanced learning, seamless learning while he continued to search ways of nurturing student interest in reading, writing and mathematics. At the beginning of this decade, realizing the need of a theory that can guide the design of future education in Asia, he worked with a small group Asian researchers, which grew bigger and an IDC Initiative was formed later, to build a learning design theory called Interest-Driven Creator (IDC) Theory. This theory may exert a far-reaching impact in future Asian education because the considerably examination-driven Asian education needs to change.

Chan bears a personal mission to facilitate the building of regional research communities since the early nineties. He co-founded two conference series, ICCE and GCCCE, and, respectively, two corresponding international academic societies, APSCE and GCSCE, one for the Asia-Pacific regional community and the other for the global Chinese community.
Dr. H. Ulrich Hoppe holds a full professorship in the area of “Collaborative and Learning Support Systems” at the University of Duisburg-Essen (Germany). After his PhD on interactive programming in mathematics education in 1984, Ulrich Hoppe has worked for about ten years in the area of intelligent user interfaces and cognitive models in HCI, before he re-focused his research on intelligent support in educational systems and distributed collaborative environments in 1995. With his COLLIDE Research Group he has participated in more than ten European projects on Technology-enhanced learning. He was one of the initiators of the European Network of Excellence Kaleidoscope (2004-07). Since 2015 he has been engaged as a PI in a Research Training Group on “User Centred Social Media” funded by the German National Science Foundation (DFG). His current research is focused on computational techniques for learning and knowledge building in various contexts, including higher education as well as vocational education and training. He is an active member of the Learning Analytics community where he particularly pursues the integration and adaptation of computational methods, such as combinations of (social) network analysis with other methods of data mining and artefact analysis.
Chee-Kit LOOI

Professor, National Institute of Education
Nanyang Technological University, Singapore
Website: https://www.nie.edu.sg/profile/looi-chee-kit

Chee-Kit Looi is Professor of Education at the National Institute of Education, Nanyang Technological University (NTU) in Singapore. He was the Founding Head of the Learning Sciences Lab, the first research centre devoted to the study of the sciences of learning in the Asia-Pacific region. He is also co-Director of the Centre of Research and Development in Learning, NTU. He organized ICCE 1995 and ICCE 2005 in Singapore, and served as President of APSCE from 1997 to 2011. He is a founding member of the Global Chinese Society of Computers in Education and served as its past president.

Chee-Kit’s research in education is characterized by producing outcomes, processes or artifacts that impact practice. An early completed project involved the design of digital mathematics manipulatives which have been made available to all secondary schools in Singapore. He is the PI or co-PI of several research projects funded by the National Research Foundation, Singapore. His research work on creating routine practices of rapid collaborative learning using GroupScribbles has made significant inroads into transforming school practices in several primary and secondary schools. His research on seamless and mobile learning has made good progress toward creating a model of 1:1 computing in schools, remarkable in terms of achieving sustainability and scalability in over ten schools.

Chee-Kit was an associate editor for the JLS, and an editorial member of JCAL, iJCSCL, and IJAIED. He was a member of the Core Expert Group that developed the framework for assessing Collaborative Problem Solving in OECD PISA 2015.
Riichiro MIZOGUCHI

Fellow
Japan Advanced Institute of Science and Technology (JAIST),
Japan
Website: http://www.jaist.ac.jp/ks/english/portfolio/mizoguchi/

Riichiro Mizoguchi received Ph.D. degree from Osaka University in 1977. He had been a full professor of the Institute of Scientific and Industrial Research, Osaka University from 1990 to 2012 and a research professor of Research Center for Service Science, Japan Advanced Institute of Science and Technology (JAIST) from October, 2012 to March, 2019. He is currently Fellow of JAIST and Associate researcher, ISTC-CNR Laboratory for Applied Ontology, Trento, Italy. His research interests include Non-parametric data analyses, Knowledge-based systems, Ontology engineering and Intelligent learning support systems. Dr. Mizoguchi was President of International AI in Education Society and Asia-Pacific Society for Computers in Education from 2001 to 2003 and President of Japanese Society for Artificial Intelligence(JSAI) from 2005-2007. He received honorable mention for the Pattern Recognition Society Award in 1985, Best paper award of the Institute of Electronics, Information and Communication Engineers in 1988, 10th Anniversary Memorial Paper Award of JSAI in 1996, Best paper awards of ICCE99 and ICCE2006, Best paper award of JSAI in 2006 and 2012, and Best paper award of Japan Society for Information and Systems in Education in 2010 and 2019. He was Vice-President of SWSA (Semantic Web Science Association) and Co-Editor-in-Chief of J. of Web Semantics from 2005 to 2009 and from 2008 to 2011, respectively. He is currently an associate editor of ACM TiiS and an editorial board member of Applied Ontology.
Gautam Biswas

Cornelius Vanderbilt Professor of Engineering
Professor, Computer Science and Computer Engineering, EECS
Department
Senior Research Scientist, Institute for Software Integrated Systems (ISIS)
Vanderbilt University, Nashville, TN. USA
Website: http://dts-web1.it.vanderbilt.edu/~biswasg/

Dr. Gautam Biswas is a Cornelius Vanderbilt Professor of Engineering and Professor of Computer Science and Computer Engineering at Vanderbilt University. He conducts research in Intelligent Systems with primary interests in applying AI and Machine Learning techniques for monitoring and control of Cyber Physical systems, and developing Open Ended Learning Environments (OELEs) for K-12 STEM and Computational Thinking (CT) learning environments. His early work focused on learning by teaching environments that promoted the social and cognitive aspects of learning. The best known work in this area, the Betty's Brain system is still used to teach multiple science topics in middle classrooms. His more recent work focuses on developing learning environments, such as CTSIM (Computational Thinking using Simulation and Modeling), C2STEM (Collaborative, Computational STEM), and SPICE (Science Projects Integrating Computing and Engineering) that exploit synergies between CT and STEM concepts and practices to help middle and high school students learn by building computational models. He has also developed innovative learning analytics approaches, using multimodal data for studying students’ self-regulated learning behaviors and linking them to their performance in the learning environment.

He has published extensively, and has over 600 refereed publications, and received a number of best paper awards. In 2011, he received the NASA Aeronautics Research Mission Directorate Technology and Innovation Group Award for Vehicle Level Reasoning System and Data Mining methods to improve aircraft diagnostic and prognostic systems. In addition to being an APSCE Fellow, Prof. Biswas is also a Fellow of the IEEE and the Prognostics and Health Management (PHM) Society.
Siu Cheung KONG

Professor, Department of Mathematics and Information Technology
Director, Centre for Learning, Teaching and Technology
The Education University of Hong Kong, Hong Kong
Website: https://www.eduhk.hk/mit/staff/sckong/

Prof. KONG Siu Cheung currently is Professor of the Department of Mathematics and Information Technology (MIT); and Director of Centre for Learning, Teaching and Technology (LTTC), the Education University of Hong Kong. Prof. Kong holds a doctorate from the Department of Computer Science of the City University of Hong Kong. Prof. Kong has produced 250 academic publications in the areas of pedagogy in the digital classroom and online learning; policy on technology-transformed education and professional development of teachers for learner-centered learning; and computational thinking education. He has completed/conducted 74 research projects since joining the University (the then Hong Kong Institute of Education). Prof. Kong is at present serving as the Editor-in-Chief of the international journal Research and Practice in Technology Enhanced Learning (RPTEL) and Journal of Computers in Education (JCE). He was in the presidential roles for the Asia-Pacific Society for Computers in Education (APSCE) for six years, as the President-Elect in 2012 and 2013, the President in 2014 and 2015, and Past-President in 2016 and 2017. Prof. Kong is the Convener of Computational Thinking Education in Primary and Secondary Schools International Research Network (IRN) under World Educational Research Association (WERA) since May 2019. He also convened the WERA IRN Theory and Practice of Pedagogical Design for Learning in Digital Classrooms from December 2012 to December 2015. Prof. Kong currently is the principal investigator of an eight-year project on coding for computational thinking development.
Wenli CHEN

Associate Professor, Learning Sciences and Assessment (LSA) Academic Group, National Institute of Education (NIE) Nanyang Technological University (NTU), Singapore

Dr. Chen Wenli is an Associate Professor with the Learning Sciences and Assessment (LSA) Academic Group at the National Institute of Education (NIE), Nanyang Technological University (NTU) Singapore. She is the Programmer Leader for LSA’s PhD and EdD programme, the co-chair for NIE AI for Education working committee, and NIE Teaching and Learning Committee member.

As a learning scientist, she specializes in Computer-Supported Collaborative Learning (CSCL), learning analytics and mobile learning. Dr Chen’s school-based design-based research projects address the challenges of transforming and enhancing teaching and learning, and applying her research outcomes to impact school practices. Through routine engagement of ICT enhanced learning activities, she has been working closely with schools in Singapore to enhance students’ collaborative learning skills and higher order thinking. Currently she is focusing on the multimodal learning analytics (MLA) of technology-enhanced learning from both behavioral and neuroscience perspectives (e.g., eye tracking, EEG, and fNIRS).

Dr. Chen is a well-known researcher within and beyond Singapore. She has been the PI and co-PI for more than 20 research projects funded by National Research Foundation Singapore and Ministry of Education (MOE) Singapore. She has published 2 books and more than 80 papers on international peer-reviewed journals. She has been the keynote speaker for 20 international conferences. She has won a dozen “Best Paper Awards” in international conferences. She was presented the “Distinguished Researcher Award” by the Asia-Pacific Society for Computers in Education in 2020. She is the Expert panel member of MOE Tertiary Education Research Fund and MOE Senior Specialist Track Research Fund. She received the “Excellence in Teaching Commendation” in 2015 and 2017 from NIE and the “Nanyang Education Award” from NTU in 2016.

Dr. Chen is the Editor-in-Chief for Journal of Computers in Education (ESCI indexed), the Associate Editor for Instructional Science: An International Journal of the Learning Sciences
(SSCI indexed), Asia Pacific Journal of Education (SSCI indexed), Research and Practice in Technology Enhanced Learning, the Advisory Editor for Asia Pacific Education Review (SSCI indexed), and the editorial board member for International Journal of Computer-Supported Collaborative Learning (SSCI indexed).

She has made significant contributions to APSCE and other international societies. Dr. Chen is currently the executive committee member for Asia Pacific Society of Computers in Education (APSCE), the CSCL community committee co-chair of International Society of the learning Sciences (ISLS) and the executive committee member of Global Chinese Society of Computers in Education (GCSCE). She was the Programme Committee Chair for International Conference on Computers in Education (ICCE) 2017, Co-Chair for the International Conference of the Learning Sciences (ICLS) 2016, the Programme Committee Chair for Global Chinese Conference on Computers in Education 2014, and the Organizing Committee Chair for International Conference on Computers in Education (ICCE) 2012.
Kaushal Kumar BHAGAT

Assistant Professor, Centre for Educational Technology
Indian Institute of Technology (IIT), India

Dr. Kaushal Kumar Bhagat is currently working as an assistant professor in the Centre for Educational Technology at the Indian Institute of Technology (IIT), Kharagpur, India. He received his Ph.D. from the Graduate Institute of Science Education at the National Taiwan Normal University in September 2016. He then served a two-year postdoctoral position at the Smart Learning Institute at Beijing Normal University. In 2015, Dr. Bhagat received NTNU International Outstanding Achievement Award. He was also awarded 2017 IEEE TCLT Young Researcher award. He is an associate editor of British Journal of Educational Technology (BJET) and editor-in-chief of Contemporary Educational Technology (CET). He is also a consultant for the Commonwealth of Learning, Canada. His research area of interest includes online learning, augmented reality, virtual reality, flipped classroom, formative assessment and technology-enhanced learning. He has published several referred journal articles and book chapters.
LAST TEN YEARS’
DISTINGUISHED RESEARCHER AWARD WINNERS

2015 - APSCE Distinguished Researcher Award
Lung-Hsiang WONG, Nanyang Technological University, Singapore

2014 - APSCE Distinguished Researcher Award
Hiroaki OGATA, Kyushu University, Japan

2011 - APSCE Distinguished Researcher Award
Antonija MITROVIC, University of Canterbury, New Zealand
Chen-Chung LIU, National Central University, Taiwan
LAST FIVE YEARS’
EARLY CAREER RESEARCH AWARD WINNERS

2019 - APSCE Early Career Researcher Award
Cheng-Jiu YIN, Kobe University, Japan

2018 - APSCE Early Career Researcher Award
Ting-Chia HSU, National Taiwan Normal University, Taiwan

2017 - APSCE Early Career Researcher Award
Jon MASON, Charles Darwin University, Australia

2015 - APSCE Early Career Researcher Award
Morris Siu-yung JONG, The Chinese University of Hong Kong, Hong Kong
Webinar 1: “Video Lecture on Affect Sequences and Learning in Betty’s Brain”
Speaker: Alexandra L. ANDRES, University of Pennsylvania, USA
Date: 9 July 2019
Curated by: APSCE Artificial Intelligence in Education SIG (AI-ED)

Webinar 2: “From Word Representation to Domain Generation”
Speaker: Yuchen LIAO, Rufu Tech., USA
Date: 17 October 2019
Curated by: APSCE Artificial Intelligence in Education SIG (AI-ED)

Webinar 3: “The Distributional Hypothesis”
Speakers: Jose Isidro BERAQUIT, Jonathan DL CASANO, & Ma. Mercedes T. RODRIGO, Ateneo de Manila University and Xiangen HU, University of Memphis
Date: 8 November 2019
Curated by: APSCE Artificial Intelligence in Education SIG (AI-ED)

Webinar 4: “Matrix Factorization and Distributional Clustering”
Speakers: Jose Isidro BERAQUIT, Jonathan DL CASANO, & Ma. Mercedes T. RODRIGO, Ateneo de Manila University
Date: 13 November 2019
Curated by: APSCE Artificial Intelligence in Education SIG (AI-ED)

Webinar 5: “Investigating the Effect of Voluntary Use of an Intelligent Tutoring System on Students’ Learning”
Speaker: Antonija (Tanja) MITROVIC, Canterbury University, New Zealand
Date: 11 December 2019
Curated by: APSCE Webinar Taskforce

Webinar 6: “Nurturing Socially Responsible Behavior with the SORBET Project – A COVID-19 Response”
Speaker: Kenneth LIM, Nanyang Technological University, Singapore
Date: 24 September 2020
Curated by: APSCE Webinar Taskforce

Webinar 7: “Building Communities of Practice in Online Education: A Designer’s Perspective”
Speaker: Sameer S SAHASRABUDHE, Educational Multimedia Research Center, India
Date: 9 November 2020
Curated by: APSCE Practice-Driven Research, Teachers’ Professional Development & Policy of ICT in Education SIG (PTP)
Peter GOODYEAR

University of Sydney, Australia

Title:
Computers in Education: Reflections on Learning, Technology and Design

Abstract: Almost exactly 40 years ago, I started work as a researcher and teacher specialising in the field of ‘computers in education’. In November 1980, I took up a position at the University of Aston in Birmingham (England). I was moving from my home disciplines of geography, environmental sciences and development studies into an educational research department. During my studies in the 1970s, I had picked up some ‘transferable skills’ – in statistics, computer programming and computer-aided data analysis. These got me jobs in university research departments in political science, geography and finally education. By early 1981, I was teaching postgraduate students to program computer assisted learning packages in the authoring language PILOT and I was carrying out research into the use of computer programs to help hearing-impaired children learn to read. In 1982, I moved back to London and spent four years working with primary school and high school teachers whose schools were taking delivery of their very first microcomputers. I became interested in Seymour Papert’s work with Logo. His 1980 book, Mindstorms: children, computers and powerful ideas, was an inspiration. I went on to write my own book, for school teachers, on Logo. It’s still my best-selling work.

Much has changed in the last 40 years, but some things have remained constant – or, one might say, have changed on the surface but remain essentially the same at a deeper level.

I’d like to use this fortuitous 40 year window to make some contrasts, and draw attention to some commonalities, between the field of ‘computers in education’ then and now.

I’ll do this by reflecting on three broad areas of change, concerning learning, technology and design. In relation to learning, I will focus on the emergence of richer views of what is involved in complex human capabilities, using examples from my work with Lina Markauskaite on epistemic fluency. In relation to technology, I want to make some observations about corollaries for research of the shift from esoteric to near ubiquitous networked devices. But mainly I will focus on design. I will address some significant shifts in conceptions of design-for-learning, with illustrations from Activity-Centred Analysis and Design (ACAD). I will share, and draw some
implications from, recent research on the ways university students reconfigure what has been
designed for their learning. I will also consider the question, raised by Ezio Manzini, of what it
means to be an ‘expert designer’ in a time when ‘everyone designs’.

Biography: Peter Goodyear is Professor of Education at The University of Sydney – a position he
took up in 2003. From 2015 to 2017 he was the founding co-director of the University’s Centre
for Research on Learning and Innovation (CRLI). Previously, he set up and co-directed the Centre
for Research on Computer-Supported Learning and Cognition (CoCo) and led the Sciences and
Technologies of Learning research network.

Peter’s research interests include design for learning, networked learning, complex learning
spaces, the nature of professional knowledge and professional education. He has published 13
books and 140 journal articles and book chapters.
Vania DIMITROVA

University of Leeds, UK

Title:
Intelligent Mentoring Systems: Tapping into AI to deliver the next generation of digital learning

Abstract: Mentoring is seen as a highly effective method to support the development of transferable skills, to increase motivation and confidence, and to develop self-regulation and self-determination. However, mentoring does not scale and can be costly – while ‘everyone needs a mentor’ not everyone can have a mentor. With the abundance of digital content and opportunities to link our educational systems with our everyday experience, there is an opportunity to develop intelligent mentors. This talk will draw a research agenda for intelligent mentoring systems, considering both the pedagogical and technological rationale. I will argue that time is ripe for building on the recent advances in AI to provide a new generation of digital learning systems capable of providing mentor-like behaviour. They would require multi-faceted ‘learner sensing’ mechanisms to get sufficient understanding of learners’ engagement and motivation by analysing the various digital traces left by the learners. Furthermore, intelligent mentors will embed strategies for promoting reflection and self-awareness through ‘personalised nudges’. I will illustrate this vision with example prototypes and will show how they can be embedded in the next generation of digital learning systems.

Biography: Vania’s background is in Mathematics (BS, MSc, Bulgaria) and applied Artificial Intelligence (PhD, Leeds University). She lead research activity on user-aware intelligent systems, focusing on knowledge-enriched user modelling and adaptation, knowledge capture, ontological modelling and information exploration. Vanya chaired the premier international conferences on user modelling (UMAP) and intelligent learning environments (AIED, ECTEL), as well as a series of international workshops on key topics related to user modelling, social systems, intelligent exploration. She was an associate editor of IEEE Transactions on Learning Technologies (IEEE-TLT) and is currently a member of the editorial boards for the personalisation journal (UMUAI) and the Int. Journal on AI in Education (IJAIED). Vanya is regularly invited as a reviewer for the European Commission ICT programme, and has served on the advisory board for the UK programme in technology-enhanced learning.
Abstract: For too long we have treated learning as an act of content consumption, not value creation. Research in the learning sciences, especially when focused on thriving in life, takes more than content and desire. We know people learn when interested, curious, passionate, and engaged, and when they feel safe, inspired, valued and validated. Yet, most growth platforms are designed to transmit “expert” content into isolated individuals, as if growth lives within the ideas and not the efforts of the growee. Instead, our work has been focused on building a connected growth platform that privileges what people are able to do with what they are learning, and not simply what they remember while learning. In this presentation, I will share alternative grounding assumptions for cultivating the type of learning in which members see themselves as capable and desirous of using that which they are learning to achieve goals they value. Here, the focus is on value creation and on the importance of connection, inspiration, and, especially in these disconnected times, the value of belonging in establishing conditions where people learn in ways that are personally transformative.

Biography: Sasha BARAB is an internationally recognized Learning Scientist who researchers, designs, and publishes on the power of immersive games and platform technologies for supporting connected growth. He is a professor in the School for the Future of Innovation in Society and the Mary Lou Fulton Teachers College, and the Executive Director of the Center for Games and Impact at ASU. His research has resulted in numerous grants, over 100 published manuscripts, and multiple game-infused innovations that have been engaged by over 250,000 players to support learning and transformation.

His current work extends the design boundaries from the bits and bytes of the game world to complex real-world ecosystems with the goal of helping all learners thrive in a complex, rapidly changing, digitally connected world. One recent project, ThriveCast, is based on an invite, enable, and release learning methodology and focused on cultivating growth and impact journeys so that more people can realize their potential. Across all work is a sensitivity to factors such as ecosystem integration, stakeholder alignment, enacted agency, and achieving sustainable and scalable outcomes.
Lung-Hsiang WONG

Nanyang Technological University, Singapore

Title:
Towards cross-fertilisation between diversified research orientations in technology-enhanced learning: Case studies of mobile learning

Abstract: After decades of advancements, the field of educational technology has witnessed scholars from diversity backgrounds (such as technology-focused, learning theory-driven, and practice-oriented) making remarkable contributions to the research and practice in their own ways. In my talk, I will unpack the aforementioned research orientations in the lens of the Technological, Pedagogical and Content Knowledge (TPACK) Framework, and explore and compare their typical sources of inspirations, research objectives and research methodology. Using my two mobile-assisted language learning projects “Chinese-PP” and “MyCLOUD” as case studies, I will discuss how educational technology researchers of different orientations would achieve reciprocity and synergy in their joint scholarly ventures.

Biography: Lung-Hsiang WONG received his PhD (major in Computer Engineering) in 1998 and is now a Senior Research Scientist and the Co-Program Director of “Learning Sciences and Innovations” at the Centre for Research in Pedagogy and Practice, National Institute of Education, Nanyang Technological University, Singapore. He became a professional researcher in the field of technology-enhanced learning since 2005.

Armed with his vast experience in conducting school-based intervention studies, he developed his niches in both mobile seamless learning and technology-enhanced language learning. He has become one of the key researchers in the mobile seamless learning field by developing the “10 Dimensions of Mobile Seamless Learning” (10D-MSL) framework which is one of the most well-referenced frameworks of the field.

He went on to spearhead a series of funded research studies on mobile and seamless learning including “MyCLOUD”, a 1:1 mobile seamless language learning model which has later been integrated to the formal Chinese curriculum in five schools. He published two books which were touted as the first two scholarly books on mobile seamless learning, namely, “Seamless Learning in the Age of Mobile Connectivity” (as the lead editor) and “Move, Language Learning! – Discovering Seamless and Mobile-Assisted Language Learning” (as the lead author). He was named the APSCE Distinguished Researcher Award in 2015 and has won nine Best/Outstanding Paper Awards at various conferences. He is the immediate Past-President of the Asia-Pacific
Society for Computers in Education and a Managing Editor of the SSCI-indexed Asia-Pacific Journal of Education.
THEME-BASED INVITED SPEAKERS

Chengjiu YIN
Kobe University, Japan
Title:
e-Books reading log based Learning Analytics

Abstract: In recent years, Learning Analytics has become an important issue in the field of computer technology. Learning Analytics is a central concern of educational institutions, as its value becomes increasingly visible (Chen, Yin, Isaías, Psotka, 2020). Different roles can get different benefits from Learning Analytics. In Japan, e-books are continually being introduced to educational institutions, and many researchers focus on doing Learning Analytics through e-books reading logs data. By using e-book systems, we can collect students’ learning behaviour logs, which recorded such behaviours as “open learning content,” “turning to the next page,” “returning to a previous page,” “adding a bookmark,” “adding a marker,” “writing a memo,” and so on.

The talk begins by reviewing the previous researches about learning analytics in the last 10 years. I then present two case studies, which are about analysing the e-book reading logs. I will introduce the data collection procedures and how the learning strategies were identified with these two case studies. The first is “Identifying Learning Strategies Using Clustering” (Yin et al., 2018). In order to identify learning strategies from the learning logs, we visualized the learning log in time series, and grouped the students into clusters based on their learning of some meaningful measurements. An important finding emerged from the analyses: The backtrack learning strategy was found to have merit as it can help students save time when studying. The second is “Examining Learning Strategies Using Sequential Analysis” (Yin et al., 2017). In order to explore the learning strategies students adopted when reading academic papers. Progressive sequential analysis was used to infer the learning strategies of students when they were reading the academic papers. The analysis results identified many significant sequences that occurred while reading the digital textbooks. We then carried out interviews to ask the participants why they took such actions. There are also some interesting findings.
**Ting-Chia HSU**

National Taiwan Normal University  
**Title:** Mobile-Assisted Language Learning Studies

**Abstract:** Dr. Ting-Chia HSU is currently a Distinguished Professor in the Department of Technology Application and Human Resource Development in National Taiwan Normal University. She is the Chair of The Special Interest Group (SIG) on Technology Enhanced Language Learning (TELL) in the computer education division in the Ministry of Science and Technology, Taiwan at present. She was also the chair of the TELL SIG in the Asia-Pacific Society for Computers in Education from 2018 to 2019. She is an associate editor of a SSCI journal named Frontiers in Psychology-Educational Psychology. She has published more than thirty SSCI journal paper and received multiple academic awards such as the Special Outstanding Talent Award, the Ta-You Wu Memorial Award rewarded by the Ministry of Science and Technology, and the winner of the Early Career Researcher Award 2018 in the Asia-Pacific Society for Computers in Education, and Academic Excellence Award in National Taiwan Normal University. Her research interests include computer education and educational technology. Dr. Hsu was awarded a government scholarship by the Ministry of Education for project research abroad from August to October in 2011 (i.e., A visiting scholar in National Institute of Education in Singapore). She was granted a project research abroad by the Ministry of Science and Technology from August 2019 to January 2020 (i.e., A visiting faculty in Massachusetts Institute of Technology, USA).

From 2007, she has developed and evaluated some mobile-assisted language learning systems. The results found that the foreign language listening comprehension of the students using partial hidden captions was similar to the listening comprehension of the students using full captions. The hidden words avoid the students inputting information through reading text from eyes. The perception processing of ears becomes important so as to adapt the students to the features of reduced forms, assimilation, elision, and so on. On the contrary, unfamiliar vocabulary which was shown in the video captions so as to be read by the eyes of the students would assist students to recognizing the vocabulary they heard and prevent the students from confusing the vocabulary they heard with other words having similar pronunciation. Finally, she conducted an experiment to compare the effects of the identical caption-filtering and personalized caption-filtering on system usability, perceived satisfaction, enjoyment, and learning motivations. In conclusion, those studies attempted to implement personalized mobile-assisted language learning systems by providing adaptively support or learning materials.
**Emma MERCIER**

University of Illinois at Urbana Champaign

**Title:**
A micro-ecological approach to the design and implementation of CSCL in Classrooms

**Abstract:** Designing and implementing CSCL for classrooms requires that we consider the interplay between multiple classroom features (teams, tasks, technology and teachers) and consider learning as it occurs across levels (individual, group, whole class). In this talk I will draw on a multi-year design-based implementation research project, focused on supporting collaborative learning in large introductory engineering courses, to illustrate how these features and levels can be addressed.
## TABLE OF CONTENTS

### Full Papers

<table>
<thead>
<tr>
<th>Paper Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automatic Assessment of Comment Quality in Active Video Watching</td>
<td>Negar MOHAMMADHASSAN, Antonija MITROVIC, Kourosh NESHATIAN &amp; Jonathan DUNN</td>
<td>1</td>
</tr>
<tr>
<td>Characterising Video Segments to Support Learning</td>
<td>Abrar MOHAMMED &amp; Vania DIMITROVA</td>
<td>11</td>
</tr>
<tr>
<td>Attainable Range Prediction of Group Product by Aggregation of Individual Products in Group Learning with Kit-build Concept Map</td>
<td>Yusuke HAYASHI, Toshihiro NOMURA &amp; Tsukasa HIRASHIMA</td>
<td>21</td>
</tr>
<tr>
<td>Utilizing Crowdsourcing and Topic Modeling to Generate Knowledge Components for Math and Writing Problems</td>
<td>Steven MOORE, Huy NGUYEN &amp; John STAMPER</td>
<td>31</td>
</tr>
</tbody>
</table>

### Short Papers

<table>
<thead>
<tr>
<th>Paper Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary students’ readiness for learning of artificial intelligence: A case study in Beijing</td>
<td>Ching Sing CHAI, Pei-Yi LIN, Morris Siu-Yung JONG, Yun DAI &amp; Thomas CHIU</td>
<td>41</td>
</tr>
<tr>
<td>Predicting Student Success for Programming Courses in a Fully Online Learning Environment</td>
<td>Neil Arvin BRETANA, Mehdi ROBATI, Aastha RAWAT, Aashi PANDEY, Shreya KHATRI, Kritika KAUSHAL, Sidarth NAIR, Gerald CHEANG, Rhoda ABADIA</td>
<td>47</td>
</tr>
<tr>
<td>LiveDots: Real time Interactive Braille Music Translator to Integrate Blind Students into Music Classes</td>
<td>Claudia GUERRERO-GARCÍA-HERAS, Óscar DÍAZ-RIBAGORDA, Lucas DE-TORRE-BARRIO, Borja MANERO &amp; María GUIJARRO-MATA-GARCÍA</td>
<td>57</td>
</tr>
<tr>
<td>Learner Model of Knowledge Grounding in Discovery Learning</td>
<td>Takuya WADA, Yuki HAYASHI, Kazuhisa SETA &amp; Tomoko KOJIRI</td>
<td>63</td>
</tr>
<tr>
<td>Using the Community of Inquiry framework to develop an educational chatbot: lesson learned from a mobile instant messaging learning environment</td>
<td>Congyao QIN, Weijiao HUANG &amp; Khe Foon HEW</td>
<td>69</td>
</tr>
<tr>
<td>Learning by Problem-Posing as Kit-Building for Structure Understanding of Polynomial Factorization</td>
<td>Sho YAMAMOTO, Hiroyoshi ENOMOTO, Yusuke HAYASHI &amp; Tsukasa HIRASHIMA</td>
<td>75</td>
</tr>
<tr>
<td>Predicting end-of-session actions considering the information of learning materials</td>
<td>Daichi TAKEHARA</td>
<td>81</td>
</tr>
<tr>
<td>PAPER TITLE</td>
<td>AUTHORS</td>
<td>PAGE</td>
</tr>
<tr>
<td>----------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td><strong>Poster Papers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Using Augmented Reality (AR) in Innovating Pedagogy: Students and Psychologists’ Perspectives</td>
<td>Hazel TRAPERO, Ryan EBARDO, Jypzie CATEDRILLA, Laiza LIMPIN, Josephine DE LA CUESTA, Cecilia LEAÑO, Michelle Renee CHING</td>
<td>87</td>
</tr>
<tr>
<td>Accuracy-aware Deep Knowledge Tracing with Knowledge State Vector Loss</td>
<td>Qiushi PAN &amp; Taro TEZUKA</td>
<td>90</td>
</tr>
<tr>
<td>ITS Promoting Realization of Misguided Self-confidence in One’s Own Comprehension</td>
<td>Tomoki ABURATANI, Kazuhisa SETA, Yuki HAYASHI &amp; Mitsuru IKEDA</td>
<td>93</td>
</tr>
<tr>
<td>Designing Technology Supported Scaffolding for Fractions Learning in Primary Schools</td>
<td>Daner SUN, Lan YANG</td>
<td>96</td>
</tr>
<tr>
<td><strong>C2: ICCE Sub-Conference on Computer-supported Collaborative Learning (CSCL) and Learning Sciences</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Full Papers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unpacking students’ modelling behaviour in the Sun-Earth system: Use of digital media tool-based epistemological resources</td>
<td>Hinal KIRI, Harshit AGRAWAL, Shanize FORTE, DurgaPrasad KARNAM &amp; Sanjay CHANDRAEKHARAN</td>
<td>99</td>
</tr>
<tr>
<td>Observing Facial Muscles to Estimate the Learning State During Collaborative Learning: A Focus on the ICAP Framework</td>
<td>Yuying CAI, Shigen SHIMOJO &amp; Yugo HAYASHI</td>
<td>119</td>
</tr>
<tr>
<td>Learning Analytics of Critical Reading Activity: Reading Hayavadana during Lockdown</td>
<td>Rwitajit MAJUMDAR, Geetha BAKILAPADAVU, Reek MAJUMDER, Mei-Rong Alice CHEN, Brendan FLANAGAN &amp; Hiroaki OGATA</td>
<td>127</td>
</tr>
<tr>
<td>Using Log and Discourse Analysis to Improve Understanding of Collaborative Programming</td>
<td>Bernard YETT, Caitlin SNYDER, Ningyu ZHANG, Nicole HUTCHINS, Shitanshu MISHRA &amp; Gautam BISWAS</td>
<td>137</td>
</tr>
<tr>
<td><strong>Short Papers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computational thinking and collaborative storytelling in learning Python</td>
<td>Chien-Sing LEE, Nandakumara VEGAN, Madhavan BALAN NAIR, James Ean-Huat OOI</td>
<td>147</td>
</tr>
<tr>
<td>Investigate the Influence of Interactive Immediacy on Collaborative Knowledge Construction in Online Discussions</td>
<td>Wanyu ZHANG, Luyue LI, Lei ZHANG, Fan YE &amp; Ning MA</td>
<td>153</td>
</tr>
</tbody>
</table>
## Poster Papers

<table>
<thead>
<tr>
<th>Paper Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Towards the Design of a Robot Peer-Tutor to Help Children Learn Math Problem-Solving</td>
<td>Aaron Nol BAUTISTA, Jabin Raymond GERARDO, Harvey LALLAVE, Patrick Luigi LATORRE, Ethel ONG, Jocelynn CU, Minie Rose LAPINID &amp; Auxencia LIMJAP</td>
<td>183</td>
</tr>
<tr>
<td>Changes in the effect of concept map-based autonomous learning under different levels of self-regulation</td>
<td>Lu CHENG, Fan CHEN, Jiayu NIU, Xueying XU &amp; Ning MA</td>
<td>186</td>
</tr>
<tr>
<td>Online Collaborative Kit-Build Concept Map</td>
<td>Aryo PINANDITO, Yusuke HAYASHI &amp; Tsukasa HIRASHIMA</td>
<td>189</td>
</tr>
<tr>
<td>Proposal of Note-map for Collaborative Reading Using an E-book System</td>
<td>Juan ZHOU, Hideyuki TAKADA &amp; Chengjiu YIN</td>
<td>192</td>
</tr>
</tbody>
</table>

## Full Papers

<table>
<thead>
<tr>
<th>Paper Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extending Deep Knowledge Tracing: Inferring Interpretable Knowledge and Predicting Post-System Performance</td>
<td>Richard SCRUGGS, Ryan S. BAKER &amp; Bruce M. MCLAREN</td>
<td>195</td>
</tr>
<tr>
<td>Introducing a Mock Technique into a Learning Support System for Program Design Based on Testability</td>
<td>Masaya MURAMATSU, Yasuhiro NOGUCHI, Satoru KOGURE, Koichi YAMASHITA, Tatsuo KONISHI &amp; Yukihiro ITOH</td>
<td>205</td>
</tr>
<tr>
<td>Reimagining the conceptualisation, design and delivery of Learning Analytics</td>
<td>Aneesha BAKHARIA &amp; Linda CORRIN</td>
<td>215</td>
</tr>
<tr>
<td>A Unified Approach for Analysing Computer Science Courses: An Australian Case Study</td>
<td>Sameera JAYARATNA, Caslon CHUA, Mohammed Eunus ALI &amp; Timos SELLIS</td>
<td>225</td>
</tr>
<tr>
<td>PAPER TITLE</td>
<td>AUTHORS</td>
<td>PAGE</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Identifying Student Engagement and Performance from Reading Behaviors in Open eBook Assessment</td>
<td>Brendan FLANAGAN, Rwitajit MAJUMDAR, Kensuke TAKII, Patrick OCHEJA, Mei-Rong Alice CHEN &amp; Hiroaki OGATA</td>
<td>235</td>
</tr>
<tr>
<td>Applying Learning Analytics to Map Students’ Self-Regulated Learning Tactics in an Academic Writing Course</td>
<td>Ward PEETERS, Mohammed SAQR &amp; Olga VIBERG</td>
<td>245</td>
</tr>
<tr>
<td><strong>Short Papers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental Design of Automated Extraction for 3-Level Tutoring Support Information in Programming Exercises</td>
<td>Yasuhiro NOGUCHI, Kousei AYABE, Koichi YAMASHITA, Satoru KOGURE, Raiya YAMAMOTO, Tatsuro KONISHI &amp; Yukihiro ITOH</td>
<td>255</td>
</tr>
<tr>
<td>A Prototype Framework for a Distributed Lifelong Learner Model</td>
<td>Patrick OCHEJA, Brendan FLANAGAN, Solomon Sunday OYELERE, Louis LECAILLIEZ &amp; Hiroaki OGATA</td>
<td>261</td>
</tr>
<tr>
<td>Using Sequence Clustering to Unveil Students' Learning Strategies and Explore the Relationship with Cognitive Load</td>
<td>Ching-Hsiang KANG, Anna Y.Q. HUANG, Owen H.T. LU, Bin-Shyan JONG, Irene Y.L. CHEN &amp; Stephen J.H. YANG</td>
<td>267</td>
</tr>
<tr>
<td>Improving Classification in Imbalanced Educational Datasets using Over-sampling</td>
<td>Zeynab MOHSENI, Rafael M. MARTINS, Marcelo MILRAD &amp; Italo MASIELLO</td>
<td>278</td>
</tr>
<tr>
<td>Applying Key Concepts Extraction for Evaluating the Quality of Students' Highlights on e-Book Reading</td>
<td>Albert YANG, Irene Y.L. CHEN, Brendan FLANAGAN &amp; Hiroaki OGATA</td>
<td>284</td>
</tr>
<tr>
<td>How Does The Quality of Students' Highlights Affect Their Learning Performance in e-Book Reading</td>
<td>Albert YANG, Irene Y.L. CHEN, Brendan FLANAGAN &amp; Hiroaki OGATA</td>
<td>289</td>
</tr>
<tr>
<td>An educational system with functions of guidance and adaptive advice to support problem solving based on basic concepts of statistics</td>
<td>Kanako SUZUKI, Tatsuhiro KONISHI</td>
<td>295</td>
</tr>
<tr>
<td>OpenLA: Library for Efficient E-book Log Analysis and Accelerating Learning Analytics</td>
<td>Ryusuke MURATA, Tsubasa MINEMATSU &amp; Atsushi SHIMADA</td>
<td>301</td>
</tr>
<tr>
<td>Sequence Pattern Mining for the Identification of Reading Behavior based on SQ3R Reading Strategy</td>
<td>Owen H.T. LU, Anna Y.Q. HUANG, Che-Yu KUO, Irene Y.L. CHEN, Stephen J.H. YANG</td>
<td>307</td>
</tr>
<tr>
<td>PAPER TITLE</td>
<td>AUTHORS</td>
<td>PAGE</td>
</tr>
<tr>
<td>---------------------------------------------------------------------------</td>
<td>---------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Improving Skill for Self-Reviewing Presentation with Robot</td>
<td>Ryotaro SEYA &amp; Akihiro KASHIHARA</td>
<td>312</td>
</tr>
<tr>
<td>Middle-School Students’ Behavior Pattern and Strategy Selection in Problem Solving: A Study Based on Data from PISA 2012</td>
<td>Yang LIU, Bo YANG, Lan WU, Baoping LI &amp; Shengquan YU</td>
<td>318</td>
</tr>
<tr>
<td>A warm-up for adaptive online learning environments – the Elo rating approach for assessing the cold start problem</td>
<td>Maciej PANKIEWICZ</td>
<td>324</td>
</tr>
<tr>
<td>Design of a Self-Reflection Model in GOAL to Support Students’ Reflection</td>
<td>Huiyong LI, Rwitajit MAJUMDAR, Yuanyuan YANG &amp; Hiroaki OGATA</td>
<td>330</td>
</tr>
<tr>
<td>Automatic Feedback Models to Students Freely Written Comments</td>
<td>Jihed MAKHLOUF &amp; Tsunemori MINE</td>
<td>336</td>
</tr>
<tr>
<td>Exploring Temporal Study Patterns in eBook-based Learning</td>
<td>Gökhan AKÇAPINAR, Mohammad Nehal HASNINE, Rwitajit MAJUMDAR, Mei-Rong Alice CHEN, Brendan FLANAGAN &amp; Hiroaki OGATA</td>
<td>342</td>
</tr>
<tr>
<td><strong>Poster Papers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assessing young learners’ situational interest in an immersive virtual reality learning environment: the role of epistemic curiosity</td>
<td>Kun-Hung CHENG</td>
<td>348</td>
</tr>
<tr>
<td><strong>C4: ICCE Sub-Conference on Classroom, Ubiquitous, and Mobile Technologies Enhanced Learning (CUMTEL)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Full Papers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peer Influence, Risk Propensity and Fear of Missing Out in Sharing Misinformation on Social Media during the COVID-19 Pandemic</td>
<td>Ryan EBARDO, Josephine DE LA CUESTA, Jypzie CATEDRILLA &amp; Santoso WIBOWO</td>
<td>351</td>
</tr>
<tr>
<td>The Influence of Augmented Reality Embedding Cognitive Scaffolds on Elementary Students’ Scientific Learning</td>
<td>Xinyue JIAO, Zifeng LIU, Su CAI</td>
<td>360</td>
</tr>
<tr>
<td><strong>Short Papers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Development and Field Testing of a MALL for Filipino with a Reusable Framework for Mobile-Based Drills</td>
<td>Jenilyn L. AGAPITO, Dominique Marie Antoinette MANAHAN, Ma. Monica L. MORENO, Jose Isidro BERAQUIT, Ingrid Yvonne HERRAS, Kevin Arnel C. MORA, Johanna Marion R. TORRES, Ma. Mercedes T. RODRIGO</td>
<td>370</td>
</tr>
<tr>
<td>Enhancing Customers’ Knowledge and Decision Making using Augmented Reality</td>
<td>Wisanoo BOONRAT, Vimita VAIDYA, Nilufar BAGHAEI, Hamid SHARIFZADEH, Ali AHMED, John CASEY</td>
<td>376</td>
</tr>
<tr>
<td>PAPER TITLE</td>
<td>AUTHORS</td>
<td>PAGE</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Investigating the effect of using the social semantic tagging-based learning guidance on science learning</td>
<td>Tien-Yu HSU, Jun-Ming CHEN, Hsin-Yi LIANG</td>
<td>382</td>
</tr>
<tr>
<td>Pandemic 2020 and Education: Responding from Kiribati</td>
<td>Faisal BADAR, Jon MASON &amp; Marebu FLOOD</td>
<td>388</td>
</tr>
<tr>
<td>Design Explorations to Support Learner’s Mental Health using Wearable Device and GOAL application</td>
<td>Taisho KONDO, Huiyong LI, Yuan Yuan YANG, Rwitajit MAJUMDAR &amp; Hiroaki OGATA</td>
<td>394</td>
</tr>
</tbody>
</table>

**Poster Papers**

<table>
<thead>
<tr>
<th>PAPER TITLE</th>
<th>AUTHORS</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construction of Reasonable Accommodation by Chatbot Using SNS and Operation of an Accessibility Center</td>
<td>Noboru UENO, Hiroyuki MITSUHARA &amp; Masami SHISHIBORI</td>
<td>400</td>
</tr>
<tr>
<td>Predicting Stag and Hare Hunting Behaviors Using Hidden Markov Model</td>
<td>Rex BRINGULA &amp; Ma. Mercedes T. RODRIGO</td>
<td>403</td>
</tr>
</tbody>
</table>

**C5: ICCE Sub-Conference on Educational Gamification and Game-based Learning (EGG)**

**Full Papers**

<table>
<thead>
<tr>
<th>PAPER TITLE</th>
<th>AUTHORS</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Story Generation System Using Player’s Emotions for Review in Game-based Learning</td>
<td>Shusuke SATO, Kaoru SUMI</td>
<td>406</td>
</tr>
<tr>
<td>Investigating the Effects of Gamifying SQL-Tutor</td>
<td>Faiza TAHIR, Antonija MITROVIC &amp; Valerie SOTARDI</td>
<td>416</td>
</tr>
<tr>
<td>Design of a Game-based Intelligent Learning Environment to Remediate Fraction Addition/Subtraction Misconceptions</td>
<td>Jarl Brent OBEDOZA &amp; Raymund SISON</td>
<td>426</td>
</tr>
</tbody>
</table>

**Short Papers**

<table>
<thead>
<tr>
<th>PAPER TITLE</th>
<th>AUTHORS</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enhancing Consumer User Experience, Education and Brand Awareness through Musical Advergames</td>
<td>Forouzan FARZINNEJAD, Hadi KHEZRIAN, Mohsen KASIRI, Nilufar BAGHAEI</td>
<td>436</td>
</tr>
<tr>
<td>Engaging the families with young children in museum visits with a mixed-reality game: A case study</td>
<td>Tien-Yu HSU, Hsin-Yi LIANG, Jun-Ming CHEN</td>
<td>442</td>
</tr>
<tr>
<td>Applying Component-based Strategy to Design an Educational Board Game for Children to Learn Chinese Characters</td>
<td>Hui-Ling HSU &amp; Zhi-Hong CHEN</td>
<td>448</td>
</tr>
<tr>
<td>Game-Based Learning in Language Education: A Review of Empirical Studies from 2009 to 2018</td>
<td>Chang XU, Xian ZHANG, Xiaohan YU, Yutong LU, Chunping ZHENG</td>
<td>454</td>
</tr>
<tr>
<td>The effect of a more knowledgeable other on resilience while playing single-player puzzle video games</td>
<td>Megan PUSEY, Kok Wai WONG &amp; Natasha RAPPA</td>
<td>460</td>
</tr>
<tr>
<td>VRetorik: A VIRTUAL REALITY VIDEO GAME TO IMPROVE PUBLIC SPEAKING SKILLS</td>
<td>Enrique FUERTES, Javier NAVARRO, Jagoba MONTES, Manuel GONZALEZ, Borja MANERO</td>
<td>466</td>
</tr>
<tr>
<td>PAPER TITLE</td>
<td>AUTHORS</td>
<td>PAGE</td>
</tr>
<tr>
<td>----------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td><strong>Poster Papers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Using a Situated Speech-based Holographic Projection System to Learn the Analects of Confucius</td>
<td>Yang-Hsin FAN, Chi-Fang HUANG &amp; Zhi-Hong CHEN</td>
<td>472</td>
</tr>
<tr>
<td><strong>C6: ICCE Sub-Conference on Technology Enhanced Language Learning (TELL)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Full Papers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Improving EFL Students’ Learning Achievements and Behaviors using a Learning Analytics-based e-book System</td>
<td>Mei-Rong Alice CHEN, Rwitajit MAJUMDAR, Gwo-Jen HWANG, Yi-Hsuan LIN, Gökhan AKÇAPINAR, Brendan FLANAGAN &amp; Hiroaki OGATA</td>
<td>474</td>
</tr>
<tr>
<td>Using Computer-like Rules to Give Automatic Grammatical Written Corrective Feedback A Case Study of Structural Particles “的”, “地” and “得”</td>
<td>Si Hui LIU, Ming Ming CHIU, Zhan WANG &amp; Sze Ming LAM</td>
<td>494</td>
</tr>
<tr>
<td><strong>Short Papers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measuring Students’ Online Language Learning Engagement: Towards the Development and Validation of a Scale</td>
<td>Lin LUAN, Yan DONG &amp; Miao CAO</td>
<td>504</td>
</tr>
<tr>
<td>Evaluation of a gamified augmented reality mobile app to support English language learning among non-native speakers</td>
<td>Dan WANG, Mas Nida MD.KHAMBARI, Su Luan WONG &amp; Abu Bakar MOHAMED RAZALI</td>
<td>510</td>
</tr>
<tr>
<td>A New Technology Design for Personalized Incidental Vocabulary Learning using Lifelog Image Analysis</td>
<td>Mohammad Nehal HASNINE, Kousuke MOURI, Gokhan AKCHAPINAR, Mahmoud Mohamed Hussien AHMED &amp; Hiroshi UEDA</td>
<td>516</td>
</tr>
<tr>
<td>Exploring Online Reading Attention and Question-Generation Task Completion of English Learners at Different Proficiency Levels by Eye Tracking Method</td>
<td>Fu- Yun YU, Wen-Wen CHENG &amp; Jon-Fan HU</td>
<td>522</td>
</tr>
<tr>
<td>Exploring English Language Learners’ Conceptions of and Engagement in a Virtual Reality Learning Environment</td>
<td>Huilin WANG, Hanyong LIU, Zhihong LU &amp; Chunping ZHENG</td>
<td>528</td>
</tr>
<tr>
<td>Examining primary students’ after-class vocabulary behavioural learning patterns in user-generated learning context: a case study</td>
<td>Yanjie SONG, Hiroaki OGATA, Yin YANG &amp; Kousuke MOURI</td>
<td>534</td>
</tr>
<tr>
<td>PAPER TITLE</td>
<td>AUTHORS</td>
<td>PAGE</td>
</tr>
<tr>
<td>---------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Learning on Country: A Game-Based approach towards preserving an Australian Aboriginal Language</td>
<td>Cat KUTAY, Deborah SZAPIRO, Jaime GARCIA, William RAFFE</td>
<td>540</td>
</tr>
<tr>
<td>TELL for Indigenous Australian languages</td>
<td>Catherine BOW</td>
<td>546</td>
</tr>
<tr>
<td>Learning Vocabularies through WeChat: A Case Study of Chinese Lower-proficiency Students</td>
<td>Chunlin LEI, Yuyang CAI</td>
<td>552</td>
</tr>
<tr>
<td><strong>Poster Papers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multi-Channel CNN-BiLSTM for Chinese Grammatical Error Detection</td>
<td>Lung-Hao LEE, Yuh-Shyang WANG, Po-Chen LIN, Chih-Te HUNG &amp; Yuen-Hsien TSENG</td>
<td>558</td>
</tr>
<tr>
<td>Development of Japanese Dictogloss Learning Support Environment for Pronunciation Learning of Japanese Speech</td>
<td>Satoru KOGURE, Hiroki HAKAMATA, Yasuhiro NOGUCHI, Tatsuhiro KONISHI, Makoto KONDO &amp; Yukihiro ITOH</td>
<td>561</td>
</tr>
<tr>
<td>A Comparative Study on the Translation Quality of Specialized and General Machine Translation Outputs</td>
<td>Xiaohan YU, Guangming YU, Zhiyang LIU &amp; Chunping ZHENG</td>
<td>564</td>
</tr>
<tr>
<td>Effect of an Automated Writing Evaluation System on Students’ EFL Writing Performance</td>
<td>Limei ZHANG &amp; Zheng HUANG</td>
<td>567</td>
</tr>
<tr>
<td>Professional Learning Community’s Views on Accessibility during Emergency Remote Teaching</td>
<td>Tom GORHAM &amp; Hiroaki OGATA</td>
<td>570</td>
</tr>
<tr>
<td><strong>C7: ICCE Sub-Conference on Practice-driven Research, Teacher Professional Development and Policy of ICT in Education (PTP)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Full Papers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effects of Academic Achievement and Group Composition on Quality of Student-Generated Questions and Use Patterns of Online Procedural Prompts</td>
<td>Fu- Yun YU &amp; Wen-Wen CHENG</td>
<td>573</td>
</tr>
<tr>
<td>School and teacher level predictors for students’ information literacy in Chinese rural and urban education</td>
<td>Liqin YU, Di WU, Sha ZHU, Feixiong CHEN, Kaige ZHU</td>
<td>582</td>
</tr>
<tr>
<td>Levels of academic teachers digital competence: Polish case-study</td>
<td>Nataliia DEMESHKANT, Katarzyna POTYRAŁA &amp; Łukasz TOMCZYK</td>
<td>591</td>
</tr>
<tr>
<td>Learning Dialogues orchestrated with BookRoll: A Case Study of Undergraduate Physics Class During COVID-19 Lockdown</td>
<td>Vijayanandhini KANNAN, Jayakrishnan WARRIEM, Rwitajit MAJUMDAR &amp; Hiroaki OGATA</td>
<td>602</td>
</tr>
<tr>
<td>When is flipped classroom more effective and why it flops</td>
<td>Khe Foon HEW</td>
<td>612</td>
</tr>
<tr>
<td><strong>Short Papers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAPER TITLE</td>
<td>AUTHORS</td>
<td>PAGE</td>
</tr>
<tr>
<td>----------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Can Computer-Based Learning Environments Mitigate Large Class Size?</td>
<td>Ryan S. BAKER, Aishah AL YAMMAHI, Joe EL SEBAALY, Ali NADAF, Tina KAPP &amp; Seth ADJEI</td>
<td>622</td>
</tr>
<tr>
<td>Experiences with e-learning as a challenge for the effective training of future generations of teachers</td>
<td>Łukasz TOMCZYK</td>
<td>628</td>
</tr>
<tr>
<td>Supporting Group Learning Using Pen Stroke Data Analytics</td>
<td>Daichi YOSHITAKE, Brendan FLANAGAN &amp; Hiroaki OGATA</td>
<td>634</td>
</tr>
<tr>
<td>What Influence Malaysian Teachers’ Satisfaction towards the FROG Virtual Learning Environment? A Structural Equation Modelling</td>
<td>Mei Lick CHEOK, Su Luan WONG, Ahmad Fauzi MOHD AYUB &amp; Rosnaini MAHMUD</td>
<td>640</td>
</tr>
<tr>
<td>International Student Perspectives on Simulated Preservice Induction</td>
<td>Jon MASON, Nicolas GROMIK, Khalid KHAN &amp; Bing LIU</td>
<td>646</td>
</tr>
<tr>
<td>Development of Computational Thinking Concepts in Scratch Programming</td>
<td>Siu-Cheung KONG, Ming LAI &amp; Cora Ka-Yuk SIU</td>
<td>652</td>
</tr>
<tr>
<td>Learning Support System for Understanding Pointers Based on Pair of Program Visualizations and Classroom Practices</td>
<td>Koichi YAMASHITA, Keisuke SAKATA, Satoru KOGURE, Yasuhiro NOGUCHI, Tatsuhiro KONISHI &amp; Yukihiro ITOH</td>
<td>658</td>
</tr>
<tr>
<td>Distance Learning Practices: A Reflective Study</td>
<td>Ali AHMED, Craig WATTERSON, Nilufar BAGHAEI, Karsten LUNDQVIST</td>
<td>664</td>
</tr>
<tr>
<td>Using Online Learning Environments to Address Digital Literacy Competencies of Construction Management Graduates</td>
<td>Thilini JAYAWICKRAMA, Shokry ABDELAAL &amp; Rhodora ABADIA</td>
<td>670</td>
</tr>
<tr>
<td><strong>Poster Papers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-book based Learning in times of Pandemic</td>
<td>Rwitajit MAJUMDAR, Mei-Rong Alice CHEN, Brendan FLANAGAN &amp; Hiroaki OGATA</td>
<td>680</td>
</tr>
<tr>
<td>Implementation Challenges of STEM Education: from Teachers’ Perspective</td>
<td>Xiaojing WENG, Morris S. Y. JONG, Thomas K.F. CHIU</td>
<td>683</td>
</tr>
<tr>
<td>Teacher Professional Development based on the DigCompEdu Framework</td>
<td>Mina GHOMI &amp; Niels PINKWART</td>
<td>689</td>
</tr>
<tr>
<td>Bridging School-based Formal and Informal Learning Spaces: A Case Study of Advancing Interest-driven Education in Singapore Schools</td>
<td>Longkai WU, Sujin HE</td>
<td>692</td>
</tr>
</tbody>
</table>
Automatic Assessment of Comment Quality in Active Video Watching

Negar MOHAMMADHASSAN*, Antonija MITROVIC, Kourosh NESHATIAN & Jonathan DUNN
University of Canterbury, Christchurch, New Zealand
*negar.mohammadhassan@pg.canterbury.ac.nz

Abstract: Active Video Watching (AVW-Space) is an online platform for video-based learning which supports engagement via note-taking and personalized nudges. In this paper, we focus on the quality of the comments students write. We propose two schemes for assessing the quality of comments. Then, we evaluate these schemes by computing the inter-coder agreement. We also evaluate various machine learning classifiers to automate the assessment of comments. The selected cost-sensitive classifier shows that the quality of comments can be assessed with high weighted-F1 scores. This study contributes to the automation of comment quality assessment and the development of personalized educational support for engagement in video-based learning through commenting.

Keywords: Video-based Learning, Learning Analytics, Applied Machine Learning, Text Classification.

1. Introduction

Educational videos have become a prevalent medium used in online learning. Learning through watching videos allows learners to acquire skills in various fields anywhere at their own pace. However, video watching can often be a passive activity due to the lack of direct interaction between students and teachers (Yousef et al., 2014). Thus, a crucial challenge in video-based learning (VBL) is providing support for engagement (Chatti et al., 2016).

AVW-Space is an online, controlled video-watching platform, which supports engagement via note-taking. Early studies with AVW-Space showed that only some students write comments, but those who do learn significantly more (Mitrovic et al., 2017a). In previous work, we used information about students’ engagement to generate personalized nudges. Empirical studies showed that nudges are effective in supporting engagement: students who received nudges were more engaged and learnt more (Mitrovic et al., 2019). However, the initial nudges were too simplistic and focused mostly on comment writing and the types of comments.

In this paper, we present the analysis of comments quality as the primarily step for designing personalized nudges which focus on comment quality. Previous studies showed that some students write shallow comments, merely repeating what was said in the videos, without thinking deeply (Mitrovic et al., 2017b). Our goal is to develop an automatic way to assess the quality of comments as students write them, and generate personalized nudges to encourage students to write better quality comments. Towards this goal, we developed two quality schemes for comments and a Machine Learning (ML) classifier that can assess the quality of comments online. Our research questions are as follows:

1. How reliable are the quality schemes when used by human coders and machine learning models to classify comments based on their quality?
2. How does the ML model perform on data collected from different student populations and experimental settings?

In Section 2, we introduce AVW-Space and previous studies on this platform, followed by an overview of related work on the assessment of written self-reports. After describing the datasets, we introduce and validate two quality schemes for the comments. Next, we investigate the performance of various ML models in assessing the quality of the comments. Finally, we reflect on the findings and propose potential improvements and applications.
2. AVW-Space

AVW-Space is an online, controlled VBL platform which was initially developed for teaching transferable skills. To create a new instance of AVW-Space, the teacher needs to select YouTube videos and specify micro-scaffolds for learning called aspects. Aspects are tags which direct students’ attention to key points of the videos or prompt students to reflect on their knowledge and experience. Initially, students watch and comment on videos individually. To write a comment, the student can stop the video at any time, enter the text of the comment and select one of the aspects (Figure 1). In the second stage, the teacher selects the comments to be displayed anonymously to the class so that students can review and rate comments made by other students in the class.

Several studies have been conducted with an instance of AVW-Space focusing on presentation skills (Dimitrova et al., 2017; Hecking et al., 2017; Mitrovic et al., 2017a; Mitrovic et al., 2019; Taskin et al., 2019). The instance contained four tutorials on giving presentations. Students were instructed to watch tutorial videos first. The aspects for the tutorial videos are: “I am rather good at this”, “I did/saw this in the past”, “I didn’t realize I wasn’t doing it”, and “I like this point”. There were also four example videos of real presentations. The students were asked to critique those examples addressing the given aspects: “structure”, “visual aids”, “delivery” and “speech”. Early studies with AVW-Space revealed that only students who wrote and rated comments improved their knowledge, while no increase in knowledge was observed for students who passively watched videos (Mitrovic et al., 2017a).

Based on the results from the initial studies, AVW-Space was enhanced by adding simple personalized nudges to foster engagement. The nudges provide hints to students based on their commenting behavior, encouraging students to write comments and providing examples of comments made by previous students. A study with the enhanced version of AVW-Space showed that students who received nudges wrote significantly more comments compared to the students who interacted with the original version of AVW-Space (Mitrovic et al., 2019). However, there were no nudges that provided feedback on the quality of comments, which is the motivation for the research presented in this paper. Although there have been several studies on comments in AVW-Space such as predicting whether a comment will receive a high number of ratings (Dimitrova et al., 2017) and identifying the pattern of vocabularies used by different types of the learners (Hecking et al., 2017), there has been no research on assessing the quality of comments.

![Figure 1. AVW-Space commenting environment.](image-url)
3. Analysis of Written Self-reports

Different forms of written self-reports analysis have been conducted to assess students’ understanding of taught concepts. Preliminary research in this area focused on the textual analysis of students’ essays (Carroll, 2007; McNamara, Crossley, & Roscoe, 2013), which are long, structured pieces of text compared to video comments. Text analysis has also been used for assessing answers to questions asked during or after teaching sessions (Arbogast & Montfort, 2016; Prevost et al., 2013). However, answers to pre-designed questions are less open-ended than comments in VBL. With the increased use of Massive Open Online Courses (MOOCs), there have been studies on online discussions (Crossley et al., 2015; Martín-Monje, Castrillo, & Mañana-Rodríguez, 2018) which are similar to comments in terms of length and flexibility in the context, but comments in AVW-Space are not conversational because students write them individually. The differences between comments and other types of self-reports underline the need for more dedicated research on comments in VBL platforms.

Many attempts have been made to investigate video annotations using linguistic features or ontologies. The focus of linguistic analysis is on syntactic and semantic components of textual data. A study on video annotations in the CLAS note-taking environment (Risko et al., 2012) proposed a scheme to identify the level of reflections in the annotations (Joksimović et al., 2019). This study used Coh-Metrix, a computational linguistics facility (Graesser et al., 2014), to derive linguistic properties to determine the depth of students’ self-reflection. However, given the differences in educational environments, this scheme cannot be applied to other contexts. The study on predicting the social value of comments in AVW-Space (Dimitrova et al., 2017) also investigated the linguistic features of the comments using Linguistic Inquiry and Word Count (LIWC) (Tausczik & Pennebaker, 2010).

Content analysis, which clusters text into different conceptual categories, is used for comparing student’s knowledge with the educational materials (Daems et al., 2014). In DiViDu (Hulsman, Harmsen, & Fabriek, 2009), four reflection categories (Observations, Motives, Effects and Goals) were suggested for written self-reflection of students on the recorded video of their communication. Then, the content of annotations in each category was analyzed manually. This reflection scheme has inspired further research on assessment of reflective notes by automated text analysis approaches (Joksimović et al., 2018; Martín-Monje et al., 2018). A study on AVW-Space comments used domain-specific concept ratio\(^1\) to infer whether a comment is on-topic (Dimitrova et al., 2017). The pitfall of this measure is that for a comment which contains only one word from the domain ontology, the domain-specific ratio is equal to 1; such comments are treated as good comments while they do not illustrate critical thinking.

Network-text analysis (NTA) is a method in content analysis to extract relations between ontologies used by the learners. NTA models the text as a network of concepts (Popping, 2000). In JuxtaLearn (Daems et al., 2014), NTA was conducted on Science, Technology, Engineering, and Mathematics (STEM) video comments to assess students’ conceptual change. NTA was also used to uncover the contextual pattern of comments in AVW-Space made by different types of learners (Hecking et al., 2017). This study showed that students with high self-regulatory skills wrote many comments containing most of the domain vocabulary, while students with weak learning skills wrote few comments using a small subset of the domain vocabulary. Overall, the content analysis of annotations is not sufficient for the quality assessment of comments since there could be high-quality comments having words from learners’ own vocabularies.

Common approaches for automating the assessments of self-reports are rule-based, dictionary-based and machine learning methods (Ullmann, 2019). The rule-based approach utilizes rules defined by a human expert and a rule engine inferring information from the textual data. Defining rules manually and evaluating them could yet be labor-intensive tasks. In the dictionary-based approach, the frequency of words in various categories is analyzed. The accuracy of the dictionary-based approach depends on the selected dictionaries. However, machine learning approaches learn patterns from data automatically and recently have been frequently used for assessing reflective texts (Crossley et al., 2019; Liu et al., 2019).

\(^1\) The number of words from the domain ontology appearing in the comment, divided by the total number of words in the comment
4. Materials and Methods

We used the data collected from the previous AVW-Space studies conducted in a first-year engineering course at the University of Canterbury in 2017, 2018 and 2019. This course used AVW-Space as an online resource for training students on presentation skills. We also used the data from a study conducted with postgraduate (PG) students (Mitrovic et al., 2016). All four studies used the same videos and aspects. The 2018 and 2019 studies shared the identical experiment design, with control and experimental groups. In addition to the standard AVW-Space features (aspects, videos and rating categories), the experimental group received nudges which encouraged students to write more comments using a variety of aspects (Mitrovic et al., 2019). However, the 2017 study (Mitrovic et al., 2017b) and the PG study did not include nudges. Table 1 presents the number of students who wrote comments and the number of comments on tutorial/example videos for each study.

Table 1. Number of Students and Comments in the Studies

<table>
<thead>
<tr>
<th></th>
<th>2017</th>
<th>2018</th>
<th>2019</th>
<th>PG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>158</td>
<td>191</td>
<td>146</td>
<td>32</td>
</tr>
<tr>
<td>Comments on tutorial videos</td>
<td>670</td>
<td>1,144</td>
<td>1,101</td>
<td>346</td>
</tr>
<tr>
<td>Comments on example videos</td>
<td>575</td>
<td>687</td>
<td>660</td>
<td>368</td>
</tr>
</tbody>
</table>

4.1 Quality Schemes

To assess the quality of comments, it is necessary to define different categories of comments. There have been several frameworks proposed for students’ reflective writing. A study on academic reflective essays identifies different types of reflection such as personal belief, lessons learned or future intentions (Ullmann, 2017; 2019). Another framework for assessing the depth of students’ self-reflections on their performance groups the contents into observation, effect, or motivation and goal (Joksimović et al., 2019). A simulation environment for cross-cultural communications classifies the user’s textual interactions with the system into different groups, such as statements on the situation and real-world stories (Dimitrova et al., 2013). However, a new labelling scheme for comments is needed for AVW-Space due to its special nature. Thus, we initially explored comments and their aspects, and proposed two quality schemes, one for comments on tutorial videos and another for those on example videos.

Table 2 presents the categories of comments on tutorial videos, with some examples from previous AVW-Space studies. This scheme contains five categories: (1) Affirmative, negative or off-topic, (2) Repeating, (3) Critical and analytical, (4) Self-reflective and (5) Self-regulating comments. Comments in categories 1 and 2 are pedagogically undesirable since they do not convey deep thinking about the videos. However, comments in category 3 show more critical thinking about the video, as learners elaborate on the video content. In category 4, learners reflect on their previous experience in relation to the video. Finally, students indicate a high level of learning in category 5 by planning how to improve their future presentations using the ideas covered in the videos. We also designed a quality scheme for comments on example videos (Table 3), which is similar to the scheme for tutorial videos with the exception of self-regulation and self-reflection categories. The reason for excluding those two categories is that the students were instructed to critique example videos using aspects like structure, visual aids, delivery and speech.

To evaluate the proposed quality schemes, we selected 167 comments via stratified sampling (110/57 comments on tutorial/example videos respectively) from 2018 and 2019 studies. Three expert coders labelled those comments independently. We used the ordinal Krippendorff’s α (Krippendorff, 2010) to calculate inter-coder agreement, due to the number of coders and ordinal categories of the schemes. Krippendorff’s α values were 0.78 and 0.69 for tutorial/example videos, respectively. Krippendorff suggests that the lowest acceptable value of α for inter-coder agreement is 0.66 (Krippendorff, 2010). After reviewing the comments on which the coders disagreed, we clarified the definitions in the schemes for further manual classification.
Table 2. Quality Scheme for Comments on the Tutorial Videos

<table>
<thead>
<tr>
<th>Category</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Affirmative, negative, off-topic</td>
<td>Comments which are irrelevant or merely affirmative/negative with no explanation.</td>
</tr>
<tr>
<td></td>
<td>e.g. [Aspect: I did/saw this in the past] “very helpful.”</td>
</tr>
<tr>
<td>2. Repeating</td>
<td>Comments which only repeat the video content.</td>
</tr>
<tr>
<td></td>
<td>e.g. [Aspect: I like this point] “limit each slide to one key idea.”</td>
</tr>
<tr>
<td>3. Critical and analytical</td>
<td>Comments which mention points that are implicitly covered in the video, or show critical thinking on the content of the video.</td>
</tr>
<tr>
<td></td>
<td>e.g. [Aspect: I like this point] “Presentations can be boring and long whereas stories are more enjoyable and can have clear direction if formulated properly.”</td>
</tr>
<tr>
<td>4. Self-reflective</td>
<td>Comments in which the learner reflects on his/her behavior and previous experience or knowledge on giving presentations.</td>
</tr>
<tr>
<td></td>
<td>e.g. [Aspect: I saw/did this in the past] “My past speeches have had very interesting beginnings.”</td>
</tr>
<tr>
<td>5. Self-regulating</td>
<td>Comments where the learner decides what they would do to improve themselves in future.</td>
</tr>
<tr>
<td></td>
<td>e.g. [Aspect: I didn’t realize I wasn’t doing this] “I will definitely be trying to smile more throughout my next presentation.”</td>
</tr>
</tbody>
</table>

Table 3. Quality scheme for comments on the example videos

<table>
<thead>
<tr>
<th>Category</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Affirmative, negative</td>
<td>Comments which are irrelevant or merely affirmative/negative with no explanation.</td>
</tr>
<tr>
<td></td>
<td>e.g. [Aspect: Visual aids] “This was helpful.”</td>
</tr>
<tr>
<td>2. Repeating</td>
<td>Comments that list or name good/bad practices in the presentations without explaining the effects and causes of the practice.</td>
</tr>
<tr>
<td></td>
<td>e.g. [Aspect: Speech] “End on a question.”</td>
</tr>
<tr>
<td>3. Critical and analytical</td>
<td>Comments which criticize examples, explain the effect of a good/bad practice in the presentation or offer advice for improvement.</td>
</tr>
<tr>
<td></td>
<td>e.g. [Aspect: Speech] “Should give more meaning to the statistic by placing it in context.”</td>
</tr>
</tbody>
</table>

4.2 Automation of Assessment using Machine Learning Approaches

To automate the assessment of comments quality using machine learning, we needed to convert comments to numerical features. We took a word count approach (Tausczik & Pennebaker, 2010) for extracting numerical features rather than full parsing, since these comments are not always grammatically correct. Therefore, the text is converted to numbers by counting the frequencies of words (or their stems) in dictionaries. We extracted 94 features for each comment using LIWC (Tausczik & Pennebaker, 2010). LIWC dictionaries are collected from various psychological constructs such as cognitive processes. Also, LIWC can calculate linguistic elements including pronouns, verbs and their tenses. As mentioned earlier, previous studies on reflective writing in various contexts also used LIWC features (Joksimović et al., 2018; Liu et al., 2019). In addition to LIWC features, we used the domain-specific ratio and unique domain-specific ratios proposed in a previous study (Mitrovic et al., 2019) to consider the topic of comments. We also created a new binary feature indicating whether the used aspect is reflective (“I didn’t realize I was doing this”, “I am rather good at this” or “I did/saw this in the past”).

2 The number of unique words from the domain ontology appearing in the comment, divided by the total number of words in the comment
We normalized all features to values between 0 and 1. To reduce the size of the feature set, we removed LIWC features that are not meaningful in this context, such as “biological” words or “personal concerns”. We also removed LIWC summary features such as “clout” and “authentic”, which are derived from primary features. Therefore, the LIWC features we selected are as follow:

- word count, word per sentence, six-letter (or more) words, dictionary words, function words, pronouns, personal pronouns, I, we, you, she/he, they, impersonal pronouns, article, prepositions, auxiliary verbs, adverbs, conjunctions, negate, verbs, adjectives, comparative, interrogatives, number, quant, affect, positive emotions, negative emotions, social, cognitive processes, insight, causation, descriptive, tentative, certainty, differentiation, perceptual processes, see, hear, feel, ingest, drives, affiliation, achieve, power, reward, risk, focus on past (past tenses/adverbs), focus on present, focus on future, relativity, informal, swear words, netspeak, assent, non-fluencies, filler.

To train different ML models and evaluate them, we merged the comments from 2018 and 2019 studies, since the experiment design was identical in those two studies. Overall, we had 1,347/2,245 comments on example/tutorial videos, which we randomly split into the training (80%) and test (20%) sets on the student level. That is, the students whose comments are in the training set have no comment in the test set. After labelling all comments manually using the quality schemes, we noticed that the data is imbalanced, as the numbers of comments in each category are different, as illustrated in Table 4 for the training set.

Table 4. Categories Distribution in Tutorial and Example Comments in the Training Set

<table>
<thead>
<tr>
<th>Category</th>
<th>Tutorial comments</th>
<th>Example comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Comments</td>
<td>48</td>
<td>924</td>
</tr>
</tbody>
</table>

To deal with the imbalanced data, we initially set class weights inversely proportional to the frequency of classes. Then, we examined traditional classifiers, such as support vector machine and decision trees. We also applied the ordinal classification (Frank & Hall, 2001) to consider the order for the quality categories. However, these classifiers did not perform well. Therefore, we took a cost-sensitive approach to define proper costs for different misclassifications regarding educational purposes. For example, classifying a comment in category 1 as 3 is worse than misclassifying a comment in class 3 as 5, because comments in category 1 show very limited engagement and therefore must be correctly detected and supported appropriately. Thus, the cost of misclassifying a comment in category 1 as 3 should be higher. We designed the cost matrices presented in Figure 2, which still consider the order of categories by increasing costs as the misclassification distance grows. In these matrices, misclassifications for categories 1 and 5 for tutorial videos and categories 1 and 3 for example videos have higher costs since the comments in category 1 need educational support the most. At the same time, a high-quality comment does not require pedagogical interventions. After many experiments with this approach, random-forest was selected as the well-performing base classifier. We refer to these classifiers as $T_a$ and $E_a$, for comments on tutorial/example videos respectively.

![Figure 2](image-url)

Figure 2. Selected Cost Matrices for Comments on Tutorial (left) and Example Videos (right).

Table 5 reports the weighted mean of metrics for evaluating the performance of the cost-sensitive classifiers. F1-score is recommended for evaluating ML models on imbalanced data (Jeni, Cohn, & De La Torre, 2013). We also used the average cost and cost-saving to evaluate models in terms of costs. Cost-saving (equation 1) is the fraction by which the actual predictions reduce the costs in the
worst case of misclassification. We started with models E_a and T_a which were trained to identify each category from the corresponding quality scheme. That is, classifier T_a identifies comments belonging to one of the five categories in the quality scheme for the tutorial videos. The performance of classifiers T_a and E_a is not satisfactory. Therefore, we considered whether it is necessary to be able to predict each category of comments individually. For example, categories 4 and 5 of the tutorial comments show a very high reflection level; in such cases, a pedagogical intervention (in the form of a nudge) is not needed. When the student initially writes a comment belonging to one of those two categories, positive feedback would be encouraging, but providing positive feedback on each high-quality comment would not be appropriate for well-performing students. For that reason, we considered various combinations of quality categories and trained different classifiers. For comments on example videos, the only category requiring a nudge is category 1, so we grouped comments from categories 2 and 3 together; the resulting classifier E_b differentiates between two types of comments (category 1 versus the union of categories 2 and 3). For tutorial comments, we explored various groupings resulting in classifiers T_b to T_d.

\[
\text{cost}\_\text{saving} = 1 - \frac{\text{cost of predictions}}{\text{Maximum Cost}} \tag{1}
\]

As can be seen in Table 5, classifiers T_c, T_d and T_e have better performance than the initial model (T_b). Classifier T_d aligns well with the ICAP framework (Chi & Wylie, 2014): active learning (just repeating the lecture) is represented by categories 1 and 2; constructive learning (adding information that was not explicitly taught) is captured by categories 3, 4 and 5. However, having only 2 categories is not enough for capturing different behaviours and providing adequate support. Classifier T_c is similar to T_d, but it distinguishes between categories 1 and 2 to provide proper support. Classifier T_e groups comment into “off-topic/short affirmative or negative”, “commenting on the video” and “reflecting”. For comments on example videos, classifier E_b predicts whether a comment is describing the video and analyzing the strengths or weaknesses of the presentation in the video. E_b and T_e were selected as the best-performing classifiers.

Table 5. The Performance of the Models on the Test Set

<table>
<thead>
<tr>
<th>Video</th>
<th>Model</th>
<th>TPR</th>
<th>FPR</th>
<th>Precision</th>
<th>F1-score</th>
<th>Avg. Cost</th>
<th>Cost-saving</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example</td>
<td>E_a: 1, 2, 3</td>
<td>0.71</td>
<td>0.22</td>
<td>0.75</td>
<td>0.71</td>
<td>3.79</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>E_b: 1, 2+3</td>
<td>0.95</td>
<td>0.21</td>
<td>0.97</td>
<td>0.96</td>
<td>0.99</td>
<td>0.98</td>
</tr>
<tr>
<td>Tutorial</td>
<td>T_a: 1, 2, 3, 4, 5</td>
<td>0.72</td>
<td>0.18</td>
<td>0.72</td>
<td>0.68</td>
<td>3.53</td>
<td>0.877</td>
</tr>
<tr>
<td></td>
<td>T_b: 1, 2, 3, 4+5</td>
<td>0.70</td>
<td>0.18</td>
<td>0.72</td>
<td>0.64</td>
<td>4.42</td>
<td>0.873</td>
</tr>
<tr>
<td></td>
<td>T_c: 1, 2, 3+4+5</td>
<td>0.80</td>
<td>0.17</td>
<td>0.80</td>
<td>0.80</td>
<td>2.86</td>
<td>0.886</td>
</tr>
<tr>
<td></td>
<td>T_d: 1+2, 3+4+5</td>
<td>0.74</td>
<td>0.26</td>
<td>0.80</td>
<td>0.73</td>
<td>3.10</td>
<td>0.877</td>
</tr>
<tr>
<td></td>
<td>T_e: 1, 2+3, 4+5</td>
<td>0.84</td>
<td>0.15</td>
<td>0.86</td>
<td>0.84</td>
<td>2.08</td>
<td>0.881</td>
</tr>
</tbody>
</table>

Figure 3 presents the confusion matrices for the selected classifiers. Classifier T_e can identify comments in category 1 correctly. The only misclassifications in category 1 is for a comment saying “hello Jim” (Jim is the name of a character in a tutorial video). Besides, some of the very short comments in higher quality categories were misclassified as 1. For instance, “guideposts” should be classified as 2+3, but it is misclassified as 1. The F1-scores for classes 1, 2+3 and 4+5 are 0.72, 0.89 and 0.72 respectively. In classifier E_b, most of the comments in category 1 are classified correctly, but there are some misclassifications in class 2+3. The two misclassifications in category 1 are for comments that use domain-specific concepts to discuss the subject of the example rather than criticizing the presentation skills of the presenters. For example, “valid points” is misclassified as 2+3 since the comment includes “valid” and “points” which are two concepts from the domain ontology.
4.3 Generalizability of the Classifiers

We also evaluated the performance of the chosen models on unseen comments, from the 2017 and PG studies. The 2017 study was done with a similar population of students (the 2017 class of the same course), who have not received nudges. Therefore, we expected the performance of the classifiers to be similar to that on the test set. On the other hand, PG students usually have much stronger learning and metacognitive skills. Therefore, we expected that the performance of the classifiers on the PG data would be worse in comparison to the performance on the training/test set.

The performance of selected classifiers on the data from 2017 and PG studies is reported in Table 6. These classifiers perform differently for these datasets because the distributions of the combined categories are different from those in the 2018/2019 studies, as highlighted in Table 7. Postgraduate students wrote more reflective comments (categories 4 and 5) than the other groups of students. Also, the PG students made slightly more high-quality comments on example videos (categories 2 and 3) than first-year students. When looking at first-year students only, the 2017 set differs from the 2018/2019 data sets; the provision of nudges in 2018/2019 resulted in a higher number of reflective comments (Mitrovic et al., 2019).

<table>
<thead>
<tr>
<th>Data</th>
<th>Model</th>
<th>TPR</th>
<th>FPR</th>
<th>Precision</th>
<th>F1-score</th>
<th>Avg. Cost</th>
<th>Cost-saving</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017</td>
<td>Eb: 1, 2+3</td>
<td>0.93</td>
<td>0.26</td>
<td>0.96</td>
<td>0.94</td>
<td>1.54</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>Te: 1, 2+3, 4+5</td>
<td>0.72</td>
<td>0.26</td>
<td>0.81</td>
<td>0.74</td>
<td>3.58</td>
<td>0.77</td>
</tr>
<tr>
<td>PG</td>
<td>Eb: 1, 2+3</td>
<td>0.96</td>
<td>0.85</td>
<td>0.96</td>
<td>0.96</td>
<td>2.08</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>Te: 1, 2+3, 4+5</td>
<td>0.68</td>
<td>0.30</td>
<td>0.70</td>
<td>0.69</td>
<td>4.61</td>
<td>0.79</td>
</tr>
</tbody>
</table>

Table 7. Percentages of categories in different data sets

<table>
<thead>
<tr>
<th>Video</th>
<th>Categories</th>
<th>Training</th>
<th>Test</th>
<th>2017</th>
<th>PG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example</td>
<td>1</td>
<td>2.66</td>
<td>2.72</td>
<td>5.91</td>
<td>2.16</td>
</tr>
<tr>
<td></td>
<td>2+3</td>
<td>97.33</td>
<td>97.28</td>
<td>94.09</td>
<td>97.83</td>
</tr>
<tr>
<td>Tutorial</td>
<td>1</td>
<td>2.67</td>
<td>2.01</td>
<td>3.58</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>2+3</td>
<td>73.48</td>
<td>74.72</td>
<td>79.40</td>
<td>58.95</td>
</tr>
<tr>
<td></td>
<td>4+5</td>
<td>23.84</td>
<td>23.26</td>
<td>17.01</td>
<td>40.75</td>
</tr>
</tbody>
</table>

5. Conclusions and Future Work

In this research, we focused on the assessment of comment quality, as a starting point towards enhancing tailored support for engagement in AVW-Space. We proposed and evaluated two quality schemes for assessing comments in AVW-Space. We also automated the quality assessment of comments using a cost-sensitive approach. We tried combining categories to further improve the performance of the classifiers and to simplify the design of pedagogical interventions. Next, we selected the best-performing combinations as the automatic quality classifiers. The generalizability of these classifiers was also assessed by evaluating them on unseen data from two studies with different experimental setups. The performance of the classifiers was slightly lower for the 2017/PG datasets. However, there is still room for improving the performance of these classifiers for low-quality comments by trying...
other cost-functions and a deeper feature-engineering. Furthermore, since the classifiers were trained
and tested only on comments about giving presentations, the generalizability of this approach for other
domains should be also investigated in future work.

The results of this study will enable us to design personalized nudges focusing on the quality
of comments a student writes. For instance, when a student submits a comment in category 1, the system
would provide an immediate nudge to encourage the learner to be more focused on the video content
in his/her comment. When a student submits a comment in category 2 or 3, a nudge could suggest more
elaboration or self-reflection. The system should also give positive feedback to the student when he/she
writes a self-reflective or self-regulating comment.

Acknowledgments

AVW-Space has been developed in collaboration with Professor Vania Dimitrova and Dr Lydia Lau
from the University of Leeds, and Dr Amali Weerasinghe. The authors are grateful to Jay Holland for
technical support. We thank Professor Peter Gostomski and Dr Alfred Herritsch for their contribution
in conducting the studies.

References

student conceptual understanding. Proceedings of ASEE Annual Conference and Exposition.


Video annotation and analytics in CourseMapper. Smart Learning Environments, 3(1), 10.


(Eds.), Artificial Intelligence in Education (pp. 84–95). Springer International Publishing.

Completion: Success in an Educational Data Mining Massive Open Online Class. Proceedings of the 8th
International Conference on Educational Data Mining. (pp. 388-391).


Devise Interactive Personalised Nudges for Active Video Watching. Proceedings of the 25th Conference on
User Modeling, Adaptation and Personalization, (pp. 22–31). Bratislava, Slovakia: ACM.

Crowdsourcing for Evaluating a Simulated Learning Environment for Interpersonal Communication and


Characteristics at Multiple Levels of Language and Discourse. The Elementary School Journal, 115(2), 210–
229.

Learner Engagement in Active Video Watching. Proceedings of the 25th International Conference on

DiViDU: Assessing the level of student reflection on recorded consultations with simulated patients. Patient
Education and Counseling, 74(2), 142–149.

Jeni, L. A., Cohn, J. F., & De La Torre, F. (2013). Facing imbalanced data - recommendations for the use of
performance metrics. In 2013 Humania association conference on affective computing and intelligent
interaction (pp. 245-251). IEEE.
Characterising Video Segments to Support Learning

Abrar Mohammeda and Vania Dimitrova

School of Computing, University of Leeds, UK

Abstract: Videos provide opportunities for engagement and independent learning and are widely used in various learning contexts. However, there are challenges with using videos for learning, e.g. long videos can reduce the concentration span, learners may become bored, not everyone can be able to detect the main points in the video, and not all parts in a video will be relevant to the learner. To address these challenges, our research aims to develop automatic ways to generate narratives by combining short video segments and tailoring this to the learner’s needs. As a first step, this paper is proposing an original framework to characterise video segments for learning by combining video content and audience attention. The input for the framework includes the video transcripts, past user interactions with the videos, and an ontology defining the core domain concepts. The output is a set of patterns that are associated with the video segments, describing the focus topic and concepts of the segment. We have applied the framework on a dataset from user studies with the AVW space for presentation skills learning, including 49 video segments that are high attention intervals from past user interactions. The video segment characterisation provides useful insights to inform recommendations and segment combinations to support informal learning.

Keywords: Video-based learning, Video characterisation, Ontologies, Presentation skills.

1. Introduction

Videos have been widely used in various learning settings to facilitate independent learning and are becoming a key platform for digital learning (June et al., 2014; Hsin & Cigas, 2013). However, there are major challenges that affect user engagement with videos. Learners prefer watching short videos as their concentration span is reduced over time (Meseguer-Martinez et al., 2017; Risko et al., 2012). Also, video content complexity could affect the engagement with videos and may cause confusion or boredom (Mongkhonvanit et al., 2019). Consequently, learners may have to watch videos many times and may not be able to identify the most relevant key points in a video. This calls for finding new ways to identify the main points in a video and to direct learners to the corresponding parts in the video (called hereafter video segments) that elaborate specific key points. These challenges are experienced at a scale with the increase of both the amount of video footage available and the number of learners who use videos for learning. Therefore, manual solutions would not scale up - using experts to analyse the video segments and identify how they will be used by potential learners can be costly and ineffective.

This calls for computational means to automate the characterisation of video segments in order to identify what concepts are covered and whether learners will grasp these concepts. To address this challenge, we are proposing a data-driven approach inspired by crowdsourcing where the interaction of past learners with a video provides an indication of the main points noted by learners in a video. To represent the domain, we use an ontology defining the main concepts in the domain and their relationships. This also allows linking segments to support learning. Our approach is developed within a PhD project, which aims to automatically characterise video segments, identify optimal video segmentation and create narratives by combining different segments.

This paper presents the first step of our approach, addressing the following research question: How to characterise video-segments for learning using learners’ interaction and video content?

To address this research question, we will present a framework for characterising video segments which specifies the domain knowledge covered in a segment. The learner comments and the video transcript of the same segment will be linked to a domain ontology to identify the main topics and concepts to characterise the video segment. By comparing the video transcript and the learner comments, we gather further indication about the usefulness of the segments for learning. The video characterisation framework is applied on a dataset from using videos for informal learning of presentation skills (Mitrovic et al, 2016). Video segments which capture high attention intervals where
learners from a previous study have noted points in the videos and have generated comments accordingly are used. The characterisation of each video segment identifies what domain aspects have been covered which can show the suitability of the segment for learning. Furthermore, we illustrate how the ontology-based characterisation allows combining segments through aggregation and linking.

The rest of the paper is organised as follows. Section 2 positions our work in relevant literature on video characterisation and points at the key contribution. Section 3 outlines the video characterisation framework, including the (a) pedagogical underpinning, (b) preliminaries, (c) main definitions for characterising video segment by identifying domain topics, concepts, their coverage and level of abstraction, and (d) the video segment characterisation pipeline. Section 4 presents the application context and dataset, while Section 5 presents and discusses the results of applying the video characterisation framework in this context. Section 6 concludes and points at future work.

2. Related Work

Video characterisation has been a major goal of research to enhance the use of videos for different purposes including using videos for learning. Manual video characterisation (Chiu et al., 2018; Colasante et al., 2016) involves asking teachers and students to characterise and highlight video-content for learning to help find video-content quickly. Similarly, Dutta & Zisserman (2019) use manual characterisation by human annotators to describe video content. To speed up the video characterisation process, Benkada & Moccozet (2017) use a constrained tracker to choose which video frame a user should characterise, and also allow collaborative characterisation by a group of people. These approaches characterise the whole video considering different purposes, including learning. Learners’ experience (captured via student annotations) are used to characterise the videos. However, manual characterisation is laborious and does not scale.

Automatic video characterisation approaches have focused on analysing the visual content of the video. For example, multiple video files that are captured by calibrated imaging devices have been characterised based on a manual characterisation of a single image frame of one of the video files (Goldenberg et al., 2019). The video characterisation presented in (Xue et al., 2017) collects metadata about the video using an assembly of computer vision techniques (shoot boundary detection, a tensor based compact representation, player detection and tracking and optical character recognition). These automated video characterisation approaches have not been applied in learning, which would require not only object detection in the video but also linking objects to a learning domain.

Semantic video characterisation which adds meaning by linking objects from the videos to a specific domain has been proposed in several works. To promote the re-use of the learning videos, Rezazadeh Azar (2017) utilises a probabilistic graphical model that shows a set of variables and their dependencies in a directed acyclic graph to exploit the semantic relationships between domain concepts. Ashangani et al. (2016) use an ontology to add semantic meaning to video characterisation to facilitate video retrieval. Videos are automatically analysed to detect shot boundaries, to extract features, and to conduct automatic text annotation. The user’s query and video content are matched by using an ontology.

A recent review of semantic video characterisation (Sikos, 2017) points at key challenges, including the wide variety of video codecs, the lack of standardised vocabularies, the vast number of video resources, the inherent ambiguity of audio-visual contents, and the unstructured nature of user-generated content. As a way to address these limitations, Sikos recommends semi-automatic or automatic video annotation using ontologies and Linked Data combined with semantic tagging tools.

The research presented here performs semantic video characterisation. Similar to the existing approaches, it uses an ontology to represent the domain of interest and utilises natural language processing and semantic tagging. However, there are some crucial differences. While current approaches advance video retrieval, we propose semantic characterisation to support the use of video segments for learning which is underpinned by a pedagogical model - Ausubel’s subsumption theory – to allow recommendation and linking of video segments. We primarily deal with textual data (video transcript and the experience of the learners captured in their comments), which relates to the learning content introduced in the video. Finally, we provide a formal description that allows the utilisation of our approach in a broad range of domains and illustrate its application in a practical learning context.
3. Video Characterisation Framework

3.1 Pedagogical Underpinning

The main goal of our video characterisation framework is to help select and combine video segments to support learning. Therefore, we need an appropriate pedagogical underpinning to inform what to include in the video characterisation and how to combine segments to support people to learn meaningfully from material presented to them (video segments in this case). Ausubel’s subsumption theory for meaningful learning (Ausubel et al., 1968) has been selected for this purpose. According to this theory, a primary process in learning is subsumption in which new material is related to relevant ideas in the existing cognitive structures derived from learning experiences. Ausubel argued that new material should be integrated with previously presented information through comparisons and cross-referencing of new and familiar concepts. Successful adoption of the subsumption theory for meaningful learning includes using concept maps (Katagall et al., 2015; Liu et al., 2018) that allow learners to group information in related modules making the connections between modules more apparent. While concept maps provide domain structure, they do not allow reasoning and automation. Similarly to concept maps, ontologies define the main concepts and relationships in a domain. However, in addition, ontologies allow reasoning to automate the generation of instruction paths through the domain. Therefore, instead of concept maps, we use an ontology to represent the domain.

Similarly to Al-Tawil et al. (2019), who operationalises the subsumption theory to generate information exploration paths through a large knowledge graph, we explore hierarchical relationships between ontology concepts to identify how to link content. The fundamental difference is that we first need to characterise video segments by mapping them to an ontology and to do this in a way that will allow supporting the subsumption processes proposed by Ausubel. We identify focus topics and concepts of video segments using reasoning over the hierarchical links in the ontology (as presented below). This enables subsumption links (derivative, correlative, super-ordinate, and combinational subsumption) to combine video segments for meaningful learning (as illustrated in Section 5).

3.2 Preliminaries

A video segment $V$ is a video interval $[V_s, V_e]$ that has a start point $V_s$ and an end point $V_e$. A video can include several video segments. We assume that each video has an audio transcript (text) and a set of textual comments made by users who have interacted with the video. Consequently, each video segment $V$ is associated with a transcript $V_t$ representing the audio in this segment and a set of user comments $V_u$ which have come from users $U$ who have interacted with this video.

We assume that the domain is defined with an ontology $\Omega = \{C, H\}$ which includes the relevant domain concepts $C \neq \emptyset$ and a concept hierarchy $H$ linking these concepts. We use $c_i \subseteq C$ to denote that $c_i$ is a subclass of $C$. Each concept $c$ has a set of immediate sub-classes $c_i \subseteq c$ which are directly linked in the concept hierarchy. The main domain topics are the top level concepts in the concept taxonomy $T_H = \{C_1, \ldots, C_m\}$, $m > 0$, i.e. $C \in T$ where $T$ is the top of the concept hierarchy.

Using semantic tagging, we can link text to concepts in an ontology. The mentions of a concept $c$ in text $t$ accumulates all mentions of $c$ and its sub-classes, i.e. $\text{mentions}_t(c) = \sum f_i$ where $f_i$ indicates the number of times that the concept $c_i$ has been mentioned in the text $t$, for all $c_i \subseteq c$. For each video segment $V$ we identify the domain concepts mentioned in the transcript $V_t$ as $M_t(c) = \{c_1, c_2, \ldots, c_p\}$ where mentions$_t(c_i) > 0$, i.e. these concepts have been mentioned at least once in the video transcript. Similarly, for each video segment $V$ we identify the domain concepts mentioned in the user comments $V_u$ as $M_u(c) = \{c_1, c_2, \ldots, c_q\}$ where mentions$_u(c_i) > 0$, i.e. $c_i$ mentioned at least once.

3.3 Focus Topics, Focus Concepts and Coverage

Based on the preliminaries, we can define focus topics and concepts presented in the video segment transcript and the user comments. A focus topic in the transcript $F_t(c)$ of a video segment $V$ is a top concept $C \subseteq T$ in the ontology that has a notable number of mentions in the video transcript $V_t$, which we indicate with a parameter $\theta$ (i.e. $\text{mentions}_t(C) \geq \theta$). Depending on $\theta$, a video segment can have several focus topics (for example, the application presented in the next section uses $\theta = 1/3$). In a similar
way, we define the **focus topic in the user comments** \( F_d(C) \) of a video segment by considering the mentions of the top concepts in the user comments \( V_u \). See example focus topics in tables 3 and 4.

Within a focus topic, we can identify corresponding focus concepts. A **focus concept** \( F_t(C,c) \) \textit{within the topic} \( C \) \textit{as expressed in the transcript} of a video segment \( V_t \) is identified if mentions \( V_t(c) \geq \theta \). Depending on \( \theta \), a video segment can have several focus concepts within the same focus topic; for example, the application presented in the next section uses \( \theta = 1/3 \), see example focus concepts in tables 3 and 4. Similarly, we define **focus concept** \( F_u(C,c) \) \textit{within the topic} \( C \) \textit{as expressed in the user comments} of a video segment counting the mentions of the sub-classes of the focus topic in \( V_u \).

It is important to know not only that a domain topic or concept are in the focus of a video segment, but also to identify how they have been covered which can help in deciding how to use them for learning. Using the ontology enables us to define the concept coverage. A domain concept is **covered broadly** if most of its immediate sub-classes have been mentioned, otherwise it is **covered narrowly**. We use a parameter \( \theta \) to set the threshold for defining the coverage type (the application presented in the next section uses \( \theta = 1/3 \)). Accordingly, we can define the type of coverage for both focus topics and focus concepts. Examples of focus topics and focus concepts’ coverage are shown in tables 3 and 4.

### 3.4 Video Segment Characterisation Pipeline

In this section, we describe the video segments characterisation pipeline (shown in Figure 1) which provides the main computation steps to apply the video characterisation framework.

**Input data.** The characterisation process begins with the use of predefined video-segments (this can be video intervals at different length, including whole videos when short). The input data required is start and the end time of the video segments, user comments associated with the segment (if available), video transcript text, and an ontology representing the domain.

**Data processing.** The text from the input data (video transcript and user comments) is processed in two steps - text processing (which extracts the main words in the text) and semantic tagging (which maps words/phrases to concepts in the ontology). To perform these steps, existing natural language processing tools can be used. The next section provides example tools used in a practical application.

**Video segment characterisation.** Using the ontology, the definitions provided in Section 3.3. can be applied to extract the domain knowledge represented in each video segment (focus topics, focus concepts, coverage). Section 5 shows the results of domain characterisation of video segments.

![Figure 1. Video Segment Characterisation Pipeline.](image)

### 4. Application Context and Dataset

#### 4.1 AVW User Studies Datasets

The video characterisation framework presented in Section 3 has been applied to characterise video segments used in the Active Video Watching (AVW-Space) in the context of learning to give pitch presentations (Mitrovic et al., 2016). We use past interactions from several studies with AVW-Space with undergraduate and postgraduate university students (Mitrovic et al., 2016; Dimitrova et al., 2017; Hecking et al., 2017), which used 8 YouTube videos - 4 tutorials and 4 examples of pitch presentations. We use the interaction of 460 learners (aggregated from past user studies) who made 3532 comments.

The video segments used here are the high attention intervals that indicate continuous stretches of videos where learners have noted something interesting in the video (Dimitrova et al., 2017). We use 49 video segments - 24 segments from tutorial videos and 25 segments from example videos. High
Attention intervals are derived by aggregating user comments. Aggregation of a set of comments $C_{om}$ is performed as follows: $A(C_{om}) \equiv \forall (com_i \in C_{om}) \exists (com_j \in C_{om}) [(com_i \neq com_j) \wedge \text{distance}(com_i, com_j) \leq \theta]$. The granularity of continuity is determined by $\theta$ indicating the interpolation gap between adjacent comments. To take into account the time required to start writing a comment, 5 seconds adjustment has been made to the start and end of the high attention intervals. We have chosen the segments of high attention of self-regulated learners who were most engaged and generated domain related comments (Hecking et al., 2017). Comments within these predefined video segments are used as input for the video segment characterisation, comments that are not within these segments have been discarded.

4.2 Data Processing

The processing of the video transcript and user comments is done in two steps (see Section 3.4.) Text processing (tokenisation and stop words removal) uses NLTK\(^1\). Then, the resumed words are matched with the domain ontology terms, using WordNet\(^2\) for synonym check (if there is no direct match found, synonymous of the text words will be found). We first use lemmatisation to get the root of the words, then use WordNet to get synonymous, then use the SequenceMatcher in Python\(^3\) to find the closest word in meaning to the ontology terms (similarity threshold of 0.85 which was identified with several tests). Figure 2 illustrates the data processing conducted in a short video transcript.

An existing ontology with key concepts related to delivering pitch presentations (Abolkasim, 2019) is used. It is organised in a hierarchy with four main classes: Structure (70 concepts), Visual Aid (95 concepts), Delivery (106 concepts), and Presentation Attribute (27 concepts). The ontology is represented in OWL and is available online\(^4\); example concepts are shown in tables 3-4.

Figure 2. Example Data Processing for Video Segment Characterisation Using the Video Transcript.

5. Video Segment Characterisation: Results and Discussion

5.1 Overview

By applying the video segment characterisation framework, we are able to get an overview of the domain topics and concepts in the video segments in our data set (see tables 1 and 2). Two domain topics are prevalent: Visual Aid (63% in the tutorials and 52% in the examples) and Delivery (58% in the tutorials and 52% in the examples). Therefore, this set of video segments can be used for learning these topics. Although tutorial videos refer to Structure and this was picked by learners when watching tutorials, learners did not pick Structure in the examples. This indicates that Structure would be a 'difficult topic' which learners could miss to see in examples. Indeed, structure is a challenging domain topic in similar soft skill domains, e.g. writing, argumentation, negotiation. Based on the characterisation, we can note another 'difficult topic' - Presentation Attribute - which was not noted by learners in any of the

---

1. https://www.nltk.org/
2. https://wordnet.princeton.edu/
4. Link to open: http://www.semanticweb.org/sc10ena/ontologies/2017/10/PresentationOntologyV1
video segments with example presentations (although these examples were selected as highly engaging presentations). Therefore, additional scaffolding (e.g. prompts) would be needed to draw the learner’s attention to Structure and Presentation Attributes when watching examples. Note that the main ontology topics which are Delivery (D), Presentation Attribute (P), Structure (S) and Visual Aid (V) hereafter will be mentioned in the tables with the abbreviation D, P, S and V, respectively.

Table 1. Summary of Focus Topics Identified in the 24 Segments of the Tutorial which Considers both Video Transcripts(T) and User Comments(C)

<table>
<thead>
<tr>
<th>Focus Topic in T and C</th>
<th>D</th>
<th>P</th>
<th>S</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focus Topic in T</td>
<td>38%</td>
<td>8%</td>
<td>33%</td>
<td>17%</td>
</tr>
<tr>
<td>Focus Topic in C</td>
<td>13%</td>
<td>4%</td>
<td>4%</td>
<td>46%</td>
</tr>
<tr>
<td>All Focus Topics</td>
<td>58%</td>
<td>29%</td>
<td>42%</td>
<td>63%</td>
</tr>
</tbody>
</table>

The identified focus concepts give further detail about the representation of the domain topics in the data set. Because the tutorials explicitly refer to domain concepts, many of these concepts were also picked by the learners. Our video characterisation framework assigned focus concepts for all tutorial video segments, apart from one. However, although each of the 25 segments with examples had a focus topic, only 8 are with specific focus concepts. This indicates that the learners have noticed domain topics in the examples but have not articulated specific concepts within these topics.

Table 2. Summary of the Focus Topics identified in the 25 Segments of the Example Videos which considers only on User Comments

<table>
<thead>
<tr>
<th>Focus Topic without Focus Concept(s)</th>
<th>D</th>
<th>P</th>
<th>S</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focus Topic with Focus Concept(s)</td>
<td>40%</td>
<td>28%</td>
<td>24%</td>
<td>28%</td>
</tr>
<tr>
<td>All Focus Topics</td>
<td>52%</td>
<td>4%</td>
<td>52%</td>
<td></td>
</tr>
</tbody>
</table>

5.2 Characterisation Based on User Comments

In both example and tutorial videos, we have user comments which indicate how the domain was noticed by the users when watching these videos. Tables 3 and 4 show the characterisation of the segments from example and tutorial videos, which are used as illustrations in the discussion below. The broad coverage of the topic means the segments can be used to give an overview at an abstract level of the focus topic or the focus concept. While the narrow coverage of the topic means the segment can be used to illustrate in depth the focus topic or the focus concept. We identified several patterns that indicate the usefulness of the video segments for learning, as follows.

Table 3. Example Focus Topics, Focus Concepts and Coverage in Segments from Example Videos(E)

<table>
<thead>
<tr>
<th>SegmentID</th>
<th>Focus Topics and Coverage</th>
<th>Focus Concepts and Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>E11</td>
<td>S(broad)</td>
<td>S-&gt;PresenterIntro(narrow)</td>
</tr>
<tr>
<td>E13</td>
<td>D(broad)</td>
<td>D-&gt;RhetoricalDevice(narrow)</td>
</tr>
<tr>
<td>E14</td>
<td>D(broad)</td>
<td></td>
</tr>
<tr>
<td>E31</td>
<td>D(narrow) V(broad)</td>
<td>D-&gt;AudienceEmotion(narrow)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>V-&gt;RecordedVoice(narrow)</td>
</tr>
<tr>
<td>E32</td>
<td>V(broad)</td>
<td></td>
</tr>
<tr>
<td>E33</td>
<td>V(broad)</td>
<td></td>
</tr>
<tr>
<td>E34</td>
<td>V(narrow)</td>
<td>V-&gt;VisualArtefact(narrow)</td>
</tr>
<tr>
<td>E35</td>
<td>V(narrow)</td>
<td></td>
</tr>
<tr>
<td>E41</td>
<td>V(narrow) D(broad)</td>
<td>D-&gt;BodyMotion(broad)</td>
</tr>
</tbody>
</table>
Table 4. Example Focus Topics, Focus Concepts and Coverage in Sample Segments from Tutorial Videos(T). Note the use of short terms for the concept names, e.g. StrCom=StructureComponent, VisArt=VisualArtefact, etc. Concepts full name will be used through the paper.

<table>
<thead>
<tr>
<th>SegmentID</th>
<th>Focus Topics and Coverage</th>
<th>Focus Concepts and Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>T11</td>
<td>Transcript: S(broad),V(narrow)</td>
<td>Transcript: V-&gt;Text(narrow)</td>
</tr>
<tr>
<td></td>
<td>Comments: S(broad),V(narrow)</td>
<td>Comments: V-&gt;Text(narrow)</td>
</tr>
<tr>
<td>T12</td>
<td>Transcript: S(broad),V(broad)</td>
<td>Comments: S-&gt;PreIntro(narrow)</td>
</tr>
<tr>
<td></td>
<td>Comments: S(broad),V(broad)</td>
<td>Comments: V-&gt;Text(narrow)</td>
</tr>
<tr>
<td>T21</td>
<td>Transcript: S(narrow)</td>
<td>Transcript: S-&gt;StrCom(narrow)</td>
</tr>
<tr>
<td></td>
<td>Comments: S(narrow)</td>
<td>Comments: S-&gt;StrCom(narrow)</td>
</tr>
<tr>
<td>T22</td>
<td>Transcript: S(broad),V(narrow)</td>
<td>Transcript: V-&gt;VisArt(narrow)</td>
</tr>
<tr>
<td></td>
<td>Comments: S(broad),V(narrow)</td>
<td>Comments: S-&gt;StrCom(narrow)</td>
</tr>
<tr>
<td>T26</td>
<td>Transcript: S(narrow),V(narrow)</td>
<td>Transcript: S-&gt;StrCom(narrow)</td>
</tr>
<tr>
<td></td>
<td>Comments: S(narrow)</td>
<td>Comments: D-&gt;AudEmo(broad)</td>
</tr>
<tr>
<td>T33</td>
<td>Transcript: V(narrow),S(narrow)</td>
<td>Transcript: V-&gt;Audio(narrow)</td>
</tr>
<tr>
<td></td>
<td>Comments: S(broad)</td>
<td>Transcript: V-&gt;Text(narrow)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Transcript: S-&gt;ConInt(broad)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Transcript: S-&gt;PreIntro(broad)</td>
</tr>
<tr>
<td>T42</td>
<td>Transcript: D(narrow)</td>
<td>Transcript: D-&gt;Para(narrow)</td>
</tr>
<tr>
<td></td>
<td>Comments: D(broad)</td>
<td>Comments: D-&gt;LanEle(narrow)</td>
</tr>
<tr>
<td>T43</td>
<td>Transcript: V(narrow),D(narrow)</td>
<td>Transcript: V-&gt;RecVoil(broad)</td>
</tr>
<tr>
<td></td>
<td>Comments: D(broad)</td>
<td>Transcript: V-&gt;Illust(broad)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Transcript: D-&gt;BodMot(broad)</td>
</tr>
<tr>
<td>T45</td>
<td>Transcript: D(narrow),V(broad)</td>
<td>Transcript: D-&gt;SpeEmo(narrow)</td>
</tr>
<tr>
<td></td>
<td>Comments: S(narrow)</td>
<td>Transcript: D-&gt;Para(narrow)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Comments: S-&gt;StrCom(broad)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Comments: D-&gt;Para(narrow)</td>
</tr>
</tbody>
</table>

The focus topic is broadly covered. Video segments with such pattern can be used to introduce the topic to learners in an abstract level without focusing on specific details. For instance, E14 shows an introductory example for Delivery and there is no focus on specific concepts within the focus topic. This pattern (there is a focus topic but there is no focus concept) is observed in 39% of the segments. Further 18% of the segments have a broadly covered focus topic and a focus concept within it. These segments can be used to introduce the focus topic at an abstract level but also to draw the learner’s attention to the specific focus concepts. For instance, E11 can be used for introducing Structure by focusing on Presenter Introduction.

The focus topic is narrowly covered. The segments following this pattern are useful for illustrating aspects of a focus topic. 12% of the segments have a narrowly covered focus topic and no focus concept, e.g. E35 is good to illustrate Visual Aid. Further 22% of the segments have narrowly covered focus topics and specific focus concepts, which can be useful for elaborating the focus topic. For instance, T21 can be used for elaborating Structure by focusing on Structure Component.

Having Two Focus Topics. The characterised video segments with this pattern can be recommended to the learners to show the link between topics. For instance, T33 shows links between use Visual Aid (focusing on use of Audio and Text) and Structure (focusing on ContentIntro and PresenterIntro). This pattern was observed rarely in our dataset (8.2% of the segments), which is expected given the short duration of the video segments. With longer video segments, or when applied to whole videos, this pattern can identify useful video content that can illustrate relationships between topics/concepts.

Segments not suitable to use on their own. From our observation of the characterisation patterns mentioned above, we found 46% of the example segments (e.g. E13) and 36% of the tutorial segments (e.g. T11 or T12) did not fall in any of the above patterns, and could not be used on their own. However,
they can be aggregated to create learning materials that cover the learning presented in them in a more coherent way. The aggregation of segments is discussed in Section 5.4.

5.3 Comparing the Video Transcript and Learner Comments

For tutorial videos, the segment characterisation allows comparing the learner attention (learner comments) and the video content (video transcript) to understand whether the learners have picked the points mentioned in the tutorial. Examining the 24 tutorial segments, we have identified three patterns.

**Full focus topic alignment.** This shows that the focus topic in the video has been noticed by the learners as shown in their comments. Segments characterised with this pattern will be good to illustrate the focus topic. In 8% of the tutorial segments, there is a full alignment on the focus topic and focus concept among the video transcript and the learner comments. For instance, T21 is a good segment to illustrate the topic *Structure* in depth as it is narrowly covered and there is a focus on one of its concepts-*Structure Component* and this is picked by the learners. In 8% of the tutorial video segments the focus topics aligned but the learners additionally focused on a concept which is not a focus in the transcript. For instance, in T12 both the video transcript and the user comments are focusing on *Structure* and *Visual Aid* and the learners are also focusing on specific concepts - *Presenter Introduction* and *Text*. This segment is useful to show the relationship between these concepts. In 8% of the segments, while the focus topics aligned, learners have missed the focus concepts in the transcript; hence, learners’ attention should be directed to notice the focus concepts. For instance, in T42 the transcript and the comments agreed that the segment is illustrating the topic *Delivery* but the learners’ attention would have to be directed to the specific concepts - *Paralanguage* and *Language Element*.

**Partial focus topics alignment.** The segments characterised with this pattern indicate that learners focus on topics and concepts but miss other topics and concepts mentioned in the transcript. Hence, when recommending these segments, the learner’s attention should be directed to the relationship between the focus topics and the elaboration of some focus topics. In 13% of the tutorial segments learners have missed some of the focus concepts in the transcript. For instance, in T45 the transcript and comments are focusing on the topic *Delivery* but the transcript is also relating *Delivery* to *Structure* by illustrating the effect of the *Structure Component* and *Paralanguage* on *Audience Emotion*. This learning is missed in the user comments which focus only on *Paralanguage*. There are 25% of the tutorial segments where the learners missed all the focus concepts in the transcript. For instance, in T42 the transcript is elaborating on the *Delivery* by focusing on *Paralanguage* and *LanguageElement*; however, this connection is missed in the comments which did not focus on any concept. In 17% of the tutorial segments the learners and the transcript have different focus concepts within the same focus topic. For instance, in T45 the transcript and learners are focusing on *Delivery* but the transcript is showing the relationship between this topic and the use of *StructureComponent* from *Structure*, while the learners are focusing on *Paralanguage* on *Delivery*. The discrepancies in focus concepts can indicate aspects of the video that may be missed by the learners, which should be considered when recommending the video segments (on their own or in combination with other).

**Misalignment of focus topics.** The cases characterised with this pattern, which shows that there is a clear deviation between the transcript and the comments, represent 21% of the tutorial segments. Consequently, the learning content in these segments is not articulate enough to make them useful learning materials. For instance, in T26 the transcript and the comments are illustrating specific concepts within two different topics, which indicates that this segment cannot be recommended.

5.4 Combining Video Segments

The video segment characterisation can be used to combine segments in order to provide a more effective way to use video segments for learning. Combining video segments includes aggregating adjacent segments and linking segments from different parts within one video or from different videos.

**Aggregating segments from the same video.** Individual segments can be too short or may not provide good enough coverage to be recommended on their own. Hence, when adjacent segments share focus topics/concepts, they can be aggregated in longer segments. The coverage allows us to further indicate how the aggregated segment can support learning. For instance, following the sample segment characterisation from Table 3 and Table 4, we can aggregate:
segments have a common focus concept and an ordinate relationship. In our case, this will allow linking concepts from more than one topic to give a more comprehensive understanding. For instance: Structure → Presentation (Ausubel et al., 1968). This linking can be done automatically, following the focus topics and concepts of video segments, and using the hierarchical links in the ontology.

**Linking through Derivative subsumption.** Derivative subsumption occurs when new material is learned as an illustration or example of an existing construct in the human cognitive structure. For instance: Structure → Presenter introduction → example, <T12, E11> - both segments have the same focus topic and focus concept; T12 describes how presenters should introduce themselves and E11 gives an example. Visual Aid → Visual artefact → {Text, Audio, Illustration}, <T22, T33, T43> - the segments have the same focus topic - Visual Aid; T22 focuses on a higher level concept Visual Artefact, which is illustrated with its sub-classes Text and Audio (in T33) and Recorded Voice and Illustration (in T43).

**Linking through Correlative subsumption.** Correlative subsumption occurs when new material is learned as an elaboration of existing concepts within the same class. For instance: Delivery → Non-verbal communication → {Paralanguage, Bodymotion}, <T42, T45, E41> - Non-verbal Communication is elaborated with two sub-classes - Paralanguage (in T42 and T45) and Body Motion (in E41).

**Linking through Super-ordinate subsumption.** Super-ordinate learning occurs by linking several learned concepts with their super-class concept. For instance: {Language, Rhetorical Device} → Verbal communication → Delivery, <T42, E12> - T42 and E12 focus on Language and Rhetorical Devices after introducing these concepts, a link to their super-class Verbal Communication can be made.

**Linking through Combinational subsumption.** Combinational subsumption occurs when new material presents relevant links but is not subsumed through a subordinate relationship or a super-ordinate relationship. In our case, this will allow linking concepts from more than one topic to give a broader view of the domain. This enables the use of common focus concepts across segments. For instance: {Structure, Delivery} → {Structure Component, Speaker Emotion, Paralanguage}, <T21, T45> - both segments have a common focus concept Structure Component, based on which they can be linked; T21 will show concepts from the topic Structure and T45 will link these concepts to the topic Delivery.

6. Conclusion

This paper proposes a semantic approach to characterise video segments to enable their use for learning. A generic video segment characterisation framework is presented which is underpinned by Ausubel’s subsumption theory for meaningful learning and utilises an ontology to represent the domain. The framework considers both video transcripts and user comments when interacting with the videos (when available). It can be applied on any video segments and is fully independent from the segmentation. The transcript-based domain characterisation is applicable to any settings, assuming that the video content discusses the specific domain. The characterisation using past learner comments and the patterns comparing transcript and comments can be used if past learners’ interactions are available. The availability of user generated interaction data is becoming highly popular, e.g. most video sharing platforms allow user comments, MOOCs integrate user interactions, such as comments, annotations, forums, to enhance the effectiveness of videos for learning. User interactions are also captured in bespoke video-based learning platforms like the AVW-Space which was used in this paper.

The video characterisation framework is applied on a dataset of 49 video segments derived from user interactions in a video-based learning system for presentation skills training. This allowed us to characterise video segments by identifying their focus topics, focus concepts, and coverage using video transcripts and comments of a fairly large user group. Based on the domain characterisation, we derive patterns which allow recommending and combining video segments to support meaningful learning.

In future work, we conduct evaluation to examine the added pedagogical value of the patterns for recommending video segments to learners. We will automate the generation of patterns and the generation of sequences of video segments based on subsumption links. Finally, we will utilise the
video characterisation to improve video segmentation based on the identified focus topics and concepts. To ensure the generality of the approach, we will apply to another context of using videos for soft skills learning, e.g. medical communication.

Acknowledgements. The authors wish to thank Prof. Tanja Mitrovic and her colleagues at the University of Canterbury New Zealand for sharing the user interaction data from past AVW-space user studies with university students at their university. The data used to derive the video segments was collected in a user study with postgraduate students using an earlier version of the AVW system at the University of Leeds, UK. The authors are grateful to Dr. Lydia Lau and Dr. Amali Weerasinge for their contribution in conducting the user study.

References

Attainable Range Prediction of Group Product by Aggregation of Individual Products in Group Learning with Kit-build Concept Map

Yusuke HAYASHI*, Toshihiro NOMURA* and Tsukasa HIRASHIMA*
*Graduate School of Advanced Science and Engineering, Hiroshima University, Japan
* hayashi@lel.hiroshima-u.ac.jp

Abstract: Recently, we can collect and analyze a variety of data in learning with the development of information and communication technology and also expect that the prediction of learning with the data enables a deep reflection for enhancing the learning experience. This paper describes a method for the attainable rage prediction of the group learning products from aggregation of an individual's concept map with the Kit-build (KB) approach. To test this method, we examined the prediction results from the data collected from a classroom lesson. The results show that most of the actual results are in good agreement with the prediction, and the comparison between the actual results and the predictions could be useful for the teacher.

Keywords: Concept map, kit-build, collaborative learning, prediction of group learning product.

1. Introduction

In recent years, the role of teachers in the classroom has changed from instructors who impart knowledge to the students to a facilitator for management of learning with the increase of the diversity of learning forms, for example in the form of lectures, collaborative learning, and problem-based learning (Carey, 1994)( Dillenbourg & Patrick, 2010)(Grabinger & Dunlap, 1995). Expectations for teachers are not only to provide students with information to improve their understanding of the subject but also to monitor and help them in the learning process. To achieve this, teachers need to recognize the subject understanding of the class, compare it with plans, and consequently coordinate the learning activities.

With the development of information and communication technology (ICT) in recent years, we can collect and analyze a variety of data. Furthermore, it is also possible to predict the learning outcomes, and the products have enabled teachers to deeply reflect on improving the understanding of the whole class. Educational data mining facilitates a variety of techniques to analyze learning data, and learning analytics provides new insight into the consequence of learning from real data (Siemens & Baker, 2012).

Learning analytics can be defined as the "trinity" of following three methodological approaches: i) content-oriented, ii) process-oriented, and iii) network analysis (Hoppe, 2017). The content-oriented approach focuses on artifacts. The traditional way is human-interpretation and coding. Recently, there have been computational methods using text-mining techniques. The process-oriented approach is sequence analysis. Traditionally, in computer-supported collaborative learning (CSCL), many studies have analyzed the system's logfiles. Martinez-Maldonado et al. proposed and developed a multi-tabletop classroom and dashboard to support collaborative learning [6]. Network analysis focuses on the network structures, including actor-actor (social) network as well as actor-artifact networks. Matsuzawa, Oshima, Oshima, Niihara & Sakai (2011) propose a tool for exploring the network structure of collaborative learning discourses. This tool visualizes the dynamics of the network structures of students, but also students and words in discourse units.
The target of this study is the content-oriented analysis of group learning, in which students organize and reflect the understanding of what they have learned in lessons. In the 21st-century skills, creativity and innovation start with "internalize given information; beliefs/actions based on the assumption that someone else has the answer or knows the truth (Scardamalia, Bransford, Kozma & Quellmalz, 2012)." In knowledge-building processes (Stahl, 2012), students make personal beliefs, which are sometimes problematic, and create a shared understanding of collaborative knowledge through social activities.

The goal of this study is the prediction of an attainable range of product in group learning in which students collaboratively organize what they have learned. One of the situations where teachers use group learning is a part of the lesson. Group learning is useful for teachers to let learners organize what the teachers have informed to the learners. However, during group learning, it is difficult for teachers to grasp students' activities and to give feedback to them because students dynamically take actions in parallel. If teachers can predicate group learning products from the understanding of individuals at the beginning of group learning, it is helpful for them to make a plan to facilitate students' activity from the start of group learning. Although it is possible to predict the group learning products from the intermediate products in groups learning, it is very difficult to predict the products from the products at the beginning of group learning. That is the reason that we focus on the understanding of individuals at the beginning of group learning.

This study focuses on the automatic assessment of concept maps (Novak & Canas, 2006) with the kit-build (KB) method (Yamasaki, K., Fukuda, H., Hirashima T. & Funai, 2010; Hirashima, Yamasaki, Fukuda & Funai, 2015) to predicate group learning that products from the understanding of individuals for the goal of teachers' facilitation to students during group learning. The KB method requests students to reconstruct concept maps from the components of the concept maps created by a teacher. We call a concept map created by a teacher "goal map" in the KB method. This method allows students to concentrate on considering the relation between concepts and makes it possible for teachers to evaluate students' understanding by the difference from the teachers' concept map (Pairai, Wunnasri, Yoshida, Hayashi, and Hirashima, 2017). Also, in collaborative learning, comparing their kit-build concept maps each other is effective to improve their maps (Nomura, Hayashi, Suzuki, and Hirashima, 2014). Kitamura et al. reported that confirmation of understanding with kit-build concept map is more effective than confirmation with fill-in-the-blank questions (Kitamura, Hayashi, and Hirashima, 2019). Nomura et al. also analyzed the propagation of individual opinions represented in their concept maps in group learning (Nomura, Hayashi, Suzuki, and Hirashima, 2014). Hayashi, Nomura & Hirashima (2019a, 2019b) categorized the patterns of aggregation of personal concept maps as group products. In addition to that, the kit-build method can automatically assess learners' concept maps by comparison with the goal map. If each learner represents his/her understanding as a concept map with the KB method before group learning, it can be the source of the prediction of the group learning products.

Here, we propose a prediction method of group learning products from the initial understanding of the members of the group based on a concept map with the KB method (KB map). In this method, each student creates a concept map as their initial understanding of the subject. While learning as a group, students compare their concept maps and construct a kit-build concept map representing a consented understanding of what they have learned. The prediction shows the possibilities of the resulting concept maps from the concept maps of students at the beginning of the lesson. This paper also demonstrates an example of the prediction result of data from experience.
2. The Automatic Assessment of Concept Maps on Kit-build Method

This study uses the KB map for checking each learner's understanding at the beginning of group learning as well as a learning activity for learners to organize and share the information they have obtained in lessons. The characteristics of the KB map is the automatic assessment of concept maps created by learners as the comparison with the goal map created by their teachers.

In the KB map, students organize their understanding through the reconstruction of the reference concept map created by a teacher (called "goal map"), and the teacher can assess the understanding of learners with the automatic comparison between the goal map and students' concept maps. Wunnasri, Pailai, Hayashi & Hirashima (2018) confirmed the validity of the automated assessment. Students and teachers can discuss a shared understanding based on the difference between their concept maps.

The goal map is a concept map created by the teacher as information that the teacher wants to provide for learners in a lecture. In the KB map, the concept map created by the teacher is called "goal map" and the decomposed one delivered to the node and link-level (kit), which is distributed to the student. Students generate their concept maps by assembling the kit.

The KB map offers a unique framework where the goal map and student maps share the same component enabling the diagnosis of the system and extraction of the difference between them. It also helps create a superimposed map from multiple students' concept maps and facilitates the assessment of the common understanding of the students. The match rate student maps and the goal map help generate a score of comprehension for the group. Information about the matches and differences of knowledge and understanding of the students can be extracted from the superimposed maps. Further, during the process of creating concept maps, the discussion and comparison of individual concept maps among the students help in unifying the components and improve the students' understanding of the concepts.

The KB map system has been proposed as a system for realizing interaction between the teacher and the students (Sugihara, Nino, Moriyama, Moriyama, Ishida, Osada, Mizuta, Hirashima, & Funaoi, 2012). This system consists of a kit build concept map creation tool "KB Map Editor" and the evaluation and support tool "KB map analyzer." KB map editor can be used to create the student map, which is functional in the tablet terminal. As these maps represent the understanding and opinions of the individuals, they are instrumental in the discussions. The KB map analyzer promptly generates a superimposed map immediately. In addition, the diagnosis of the system makes it possible for the teacher to recognize the common understanding of the entire class.

The KB map can automatically compare the concept maps since the components of the goal map and students map are unified. The comparison may be performed for each proposition, and between students for each link, by comparing the goal map, to realize the group product prediction function. Proposition, which is determined from the above-described problems of group activities, becomes a goal map. Individual concept maps represent the understanding of each member of the group at the beginning of the learning process. The classification between the propositions, such as "the same proposition as the goal map," "divergent proposition from the goal map," and "no proposition (the link is not connected with any concepts)" is obtained by comparing the individual maps and the goal map. Then, it is possible to aggregate a combination thereof or matching rate for each group.

In the present study, the KB map analyzer processes the predictions of the products. Thus, by using the KB map analyzer, it is possible to grasp the predicted score and predict the group learning products in the form of an agreement rate for the entire class in real-time. In addition, it is possible to refer to the superimposition of individual maps of the group members for each group and to understand the particular state of the group.
3. Prediction of the group learning products with Kit-build Concept Maps

Here we employ the Kit-build concept map as the digital tool for representation and assessment of students' understanding. First, each student makes a concept map using the components provided by the teacher. The concept maps represent his/her understanding. After that, through a group discussion using the concept maps, students exchange their opinions to reach a shared agreement of the group.

We classify the proposition made by individuals in the group and pattern them. There are three types of propositions based on the link between concepts: **same** (the link connects the same concepts as the reference map), **divergent** (the link connects the divergent concepts from the reference map), and **no** (the link does not connect any concepts).

Based on the classification, Table 1 shows the possible state of personal propositions in group learning. Patterns A, B, and C indicate that all members have the same type of proposition. For example, in pattern A, all the members have the **same** proposition as the reference concept map. By contrast, in patterns D - G, there is a conflict of propositions in the group. For example, in the pattern D, some members have the **same** proposition as the reference map, but other members have **divergent** ones from the reference map.

<table>
<thead>
<tr>
<th>pattern</th>
<th>Same</th>
<th>Divergent</th>
<th>No</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>exist</td>
<td>non-existent</td>
<td>non-existent</td>
<td><strong>Same</strong></td>
</tr>
<tr>
<td>B</td>
<td>non-existent</td>
<td>exist</td>
<td>non-existent</td>
<td><strong>Divergent</strong></td>
</tr>
<tr>
<td>C</td>
<td>non-existent</td>
<td>non-existent</td>
<td>exist</td>
<td><strong>Same/Divergent/No</strong></td>
</tr>
<tr>
<td>D</td>
<td>exist</td>
<td>exist</td>
<td>non-existent</td>
<td><strong>Same/Divergent</strong></td>
</tr>
<tr>
<td>E</td>
<td>exist</td>
<td>non-existent</td>
<td>exist</td>
<td><strong>Same</strong></td>
</tr>
<tr>
<td>F</td>
<td>non-existent</td>
<td>exist</td>
<td>exist</td>
<td><strong>Divergent</strong></td>
</tr>
<tr>
<td>G</td>
<td>exist</td>
<td>exist</td>
<td>exist</td>
<td><strong>Same/Divergent</strong></td>
</tr>
</tbody>
</table>

When a group has a shared understanding of a proposition through group discussion, in which the state of the proposition moves to pattern A ~ C, there are three types of changes, i) "they select existing propositions," ii) "they generate a new proposition," and iii) "they do not make a proposition." For example, from the pattern D, if they select the existing proposition, the shared understanding of the group is "the same proposition" or "divergent proposition." However, if they generate a new proposition, they must create a **divergent** proposition as a shared understanding of their group.

We can anticipate the shared understanding of each proposition from the patterns of propositions in the group. The basic rules of prediction in this study are straightforward. If anyone in a group makes a proposition with a link, they select the proposition as their decision. This is based on the analysis of the group learning products and individual concept maps. Nomura et al. (2014) reported that learners choose the existing propositions in individual concept maps as the group learning products. However, if no one has any proposition with a link, they create the **same** or divergent proposition or do not generate any proposition. For example, in pattern A, their product is uniquely decided into the **same** proposition. However, in pattern D, they can create the **same** or divergent proposition. Besides, in patterns C, E, F, and G, they can also decide not to create any propositions.

The proposed prediction of propositions enable to calculate the prediction of the maximum and the minimum resultant map score from the proposition patterns in the group. The maximum occurs when the students in a group choose only the **same** propositions. The minimum occurs when they choose some **divergent** propositions even if a single member in the group has opted for some of the **same** propositions. In the patterns D or G, the maximum result
is derived when they choose the same proposition, and the minimum score is derived when they choose the divergent proposition. Table 2 show an example of the prediction of propositions and map score. The range of map score expected their propositions are from 80 to 20. In addition to that, their resultant map score might be 100, if they would find the same proposition as the GM about Prop. C in their group discussion. Therefore, the attainable range of map score of the group is from 20 to 100.

Table 2. An example of the prediction of propositions and the attainable range of map score

<table>
<thead>
<tr>
<th>Learner 1</th>
<th>Prop. A</th>
<th>Prop. B</th>
<th>Prop. C</th>
<th>Prop. D</th>
<th>Prop. E</th>
<th>Map score (0-100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learner 2</td>
<td>same</td>
<td>div.</td>
<td>no</td>
<td>same</td>
<td>div.</td>
<td>40</td>
</tr>
<tr>
<td>Learner 3</td>
<td>div.</td>
<td>same</td>
<td>no</td>
<td>same</td>
<td>same</td>
<td>60</td>
</tr>
<tr>
<td>Prediction</td>
<td>same/div.</td>
<td>same/div.</td>
<td>no</td>
<td>same</td>
<td>same/div.</td>
<td>100-80-20</td>
</tr>
</tbody>
</table>

same: the same proposition as the GM, div.: divergent proposition from the GM, no: no proposition

4. Experimental application to the lesson data

4.1 lesson design

This study used the data from a lesson where a total of 70 people from two classes in the second grade of the junior high school participated. These lessons were taught as one lesson for each class. The topic was the Tohoku region of Japan in geography. The purpose of this lesson was to frame together with the knowledge of nature, industry, traditions, and the culture of the Tohoku region as one structure and to make a common background for the discussion on their proposals for the reconstruction of the region in later lessons.

Figure 1 shows the basic flow of the lesson. In the first phase, like the introduction, the teacher reviewed the previous lessons about the Tohoku region with the students. In the second phase, the students created a concept map from the kit individually, and then in groups of three or four students, they combined their maps to create a group concept map. If needed, they could change their individual concept map. Finally, in the last phase, the students compared the group learning products and discussed how to make a better structure with the teacher.

In this lesson, each student used a tablet computer to create a personal map. In addition, each group had another tablet for creating a group map.

Fig. 1. the basic flow of the lesson.

Figure 2 shows the goal map and Fig. 3 shows a kit made from the goal map. Although a kit is generally made by separating all the nodes and links in the goal map, in the kit links and nodes connected to the central topic of the kit were not separated to show the the core structure of this map. The core structure represents the viewpoints to organize the the characteristics of
Tohoku region, "nature," "industry," and "culture." The tasks of the students for creating a map were to organize the instances of the viewpoints and to find intersections of the viewpoints. The students used this kit to create both a personal map and a group map.

Fig. 2. The goal map (in Japanese)

Fig. 3. The kit made from the goal map (in Japanese)

4.2 Verifications group learning product prediction

In this study, we verified the accuracy of the prediction of group activity outcomes. The accuracy is measured by the range of the expected score and its validity. The range is the
difference between the possible maximum and the minimum scores calculated from individual maps. The validity is whether the scores of the group maps fall into the predicted ranges. We used the data of the actual lesson described in the previous section.

This study represents the consistent ratio of the actual score and predicted scores in the format shown in Fig. 4 and Fig. 5. The vertical axis represents the predicted score in the interval scale. The horizontal axis represents the match rate. The maximum and minimum values on the vertical axis and the horizontal axes are the maximum and minimum of the values taken by all the groups. The solid bar indicates the minimum score from the maximum score. The dotted bar chart without color indicates the maximum possible score from the maximum score. The middle number of the bar graph indicates an ID that represents the group.

Figures 4 and 5 show the actual scores of the group learning products. The triangle indicates the actual group map score. The solid boxes represent the range of predicted score from the same and the divergent propositions in each group and the dotted boxes represent the range of predicted score if the groups generate the same propositions from no propositions in individual maps. The actual group map score in the solid box means the group has collected their same and the divergent propositions in the individual maps. If the actual group map score is out of the solid box, it means they generate the same propositions from no propositions in individual maps or from only divergent propositions in individual maps.

In Class A score prediction, the average prediction width was 23%, and the range of predicted scores were approximately in 20 points. Further, as shown in Fig. 5, seven out of nine groups are in the expected range, with an accuracy of approximately 80%. In Class B score prediction, the average prediction width was 17%, and nine out of ten groups were in the expected range; the prediction scores were approximately 20 points with 90% accuracy.

To verify the validity of the prediction, we compared the predicted and the measured values for each proposition in a pattern. Table 3 shows the number of propositions in each pattern and the percentage of propositions conforming to the prediction rules for each pattern in the total of two classes. Most of the patterns were found to be in accordance with the prediction rule. However, in pattern B and F, the ratio of the proposition conforming to the rules is low.

<table>
<thead>
<tr>
<th>Patterns</th>
<th>Same</th>
<th>Div.</th>
<th>No</th>
<th>Result of prediction</th>
<th># of total propositions</th>
<th># of successful prediction</th>
<th>Prediction Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>x</td>
<td></td>
<td></td>
<td>Same</td>
<td>26</td>
<td>26</td>
<td>1.00</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>x</td>
<td></td>
<td>Div.</td>
<td>13</td>
<td>5</td>
<td>0.36</td>
</tr>
<tr>
<td>C</td>
<td>x</td>
<td></td>
<td></td>
<td>Same/Div./No</td>
<td>49</td>
<td>49</td>
<td>1.00</td>
</tr>
<tr>
<td>D</td>
<td>x</td>
<td>x</td>
<td></td>
<td>Same/Div.</td>
<td>22</td>
<td>20</td>
<td>0.91</td>
</tr>
<tr>
<td>E</td>
<td>x</td>
<td></td>
<td></td>
<td>Same</td>
<td>81</td>
<td>73</td>
<td>0.90</td>
</tr>
<tr>
<td>F</td>
<td></td>
<td>x</td>
<td>x</td>
<td>Div.</td>
<td>70</td>
<td>29</td>
<td>0.41</td>
</tr>
<tr>
<td>G</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>Same/Div.</td>
<td>43</td>
<td>40</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Pattern B is the case where the group exclusively possesses divergent propositions. In this case, there may be several divergent propositions in the group. For example, all the members have different divergent propositions. Although the rule assumes that they adopt one of the divergent propositions, they actually have made a new divergent proposition in many cases. In 46.2% of Pattern B, the learners adopt a new divergent proposition. While this does not change the predicted score, this is different from the assumption of the rule. Furthermore, in 83.3% of the case where the learners adopt a new divergent proposition, the groups had some divergent propositions.

Pattern F is a case where divergent proposition and no proposition exist at the same time. In this case, the prediction rule assumes that the group adopts the existing divergent proposition in the group. In fact, unlike the rule, 39.2% of the groups were to build a new divergent
proposition. In addition, the "divergent proposition" in the case of 87% of which was exclusive. Even in such a case, it is considered that there is a tendency to create a new "divergent proposition" through discussion. However, to clarify the cause, further analysis is required.

![Diagram](image)

**Fig. 4.** Prediction of class A (sorted by the match rate)

![Diagram](image)

**Fig. 5.** Prediction of class B (sorted by the match rate)

### 4.3 Questionnaire and Interviews

We conducted a questionnaire and interviews with the teacher about the prediction using the kit build concept map. The contents of the questionnaire are shown in Table 4. The answers are on a scale of "Yes, I think so strongly," "Yes, I think so," "Yes and no," "No, I do not think so," and "No, I do not think so strongly" with respect to each question in the questionnaire.

In Questions 1 and 2, since the answer is "Yes, I think so," the teacher could predict the outcome of group activities to some extent from his experience in the ordinary lesson, and the prediction by the Kit-build concept map provided similar prediction. The teacher answered,
"No, I do not think so" in Question 3 and "I think so," in Question 4 on the condition that teachers are familiar with the interpretation of the graph. This suggests that the prediction can be helpful for the facilitation of the groups, although it is challenging to facilitate group learning in ordinary lessons.

The teacher also gave feedback about the prediction. The teacher considered that he could judge which groups needed help with the prediction. The teacher said that in the prediction, he wanted to check the detail of the group having a long dotted rectangle, that is, the groups have no idea about many propositions. The group with a high concordance rate of propositions is also the target of the check. It was also argued that the comparison between the actual result and the prediction is also helpful for analysis after the lesson. For example, Group 5 in Class A could have a good discussion because they have improved their map more than the potential, that is, they made some new same propositions in the group through the discussion. Meanwhile, Group 10 in Class B did not have a good discussion because their group map score is low in the predicted range; that is, they have adopted divergent propositions from the ideas in their group. He said that teachers could analyze what happens in the group learning during and after lessons if teachers could understand the graph.

Table 4. the result of the questionnaire (n = 1)

<table>
<thead>
<tr>
<th>Question</th>
<th>answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  Can you predict the group learning result in ordinary lessons?</td>
<td>2</td>
</tr>
<tr>
<td>2  Did the prediction by the Kit-build concept map provide a similar</td>
<td>2</td>
</tr>
<tr>
<td>prediction of you?</td>
<td></td>
</tr>
<tr>
<td>3  Can you consider the treatments of learning in groups?</td>
<td>4</td>
</tr>
<tr>
<td>4  Could you consider the treatments of learning in groups with the</td>
<td>2</td>
</tr>
<tr>
<td>prediction by the Kit-build concept map?</td>
<td></td>
</tr>
</tbody>
</table>

Answers:

5. Conclusion and Future Challenges

This paper proposes the attainable range prediction of group learning products by a simple aggregation of individual concept maps of group members. This is based on the automated assessment of concept maps on KBmap. KBmap provide common component of concept maps for learners and it is easy to compare concept maps made of the common component. The proposed method in this paper simply aggregates personal concept maps of group members with this mechanism and predicts group learning products.

In this case study, the score of the group activities had a 20% width, with more than 80% of the group scoring in the range of the predicted score as a result. About 80% of the proposition, even in the detailed analysis, has put the group learning products with the prediction rules. This shows the validity of the proposed prediction method in this study. In the questionnaire, the teacher pointed out the availability of the prediction graph. The group learning product prediction graph is expected to perform as the representation of the grasping ability of each group and the facilitation of group activities.

Future challenges are the verification of the use of the attainable range prediction of group learning products in the classroom by teachers and the learning effect of facilitation of group learning based on the prediction.
References

Carey, D. M. (1994) Teacher Roles and Technology Integration: Moving from Teacher as Director to Teacher as Facilitator. Computers in the Schools, 9(2), 105-118
Utilizing Crowdsourcing and Topic Modeling to Generate Knowledge Components for Math and Writing Problems

Steven MOORE*, Huy NGUYENb & John STAMPERc
abc Human-Computer Interaction Institute, Carnegie Mellon University, United States
*StevenJamesMoore@gmail.com

Abstract: Combining assessment items with their hypothesized knowledge components (KCs) is critical in acquiring fine-grained data on student performance as they work in an ed-tech system. However, creating this association is an arduous process and requires substantial instructor effort. In this study, we present the results of crowdsourcing KCs for problems in the domain of mathematics and English writing, as a first step in leveraging the crowd to expedite this task. We presented crowdworkers with a problem in each domain and asked them to provide three underlying skills required to solve it. These inputs were then analyzed through two topic modeling techniques, to compare how they might cluster around potential KCs. Results of the models’ output were evaluated against KCs generated by domain experts to determine their usability. Ultimately, we found that half of the crowdsourced KCs matched expert-generated KCs in each problem. This work demonstrates a method to leverage the crowd’s collective knowledge and topic modeling methods to facilitate the process of generating KCs for assessment items, which can be integrated in future learnersourced environments.

Keywords: Knowledge Component, Crowdsourcing, Topic Modeling, Text Mining, Knowledge Tracing, Intelligent Tutoring Systems

1. Introduction

Many educational technologies, such as intelligent tutoring systems and online courseware, utilize knowledge component modeling to support their adaptivity. This treats student knowledge as a set of interrelated KCs, where each KC is “an acquired unit of cognitive function or structure that can be inferred from performance on a set of related tasks” (Koedinger, Corbett, and Perfetti, 2012). Operationally, a KC model is defined as a mapping between each question item and a hypothesized set of associated KCs that represent the skills or knowledge needed to solve that item. This mapping is intended to capture the student’s underlying cognitive process and is vital to many core functionalities of an intelligent educational software, enabling features such as adaptive feedback and hints (Moore and Stamper, 2019).

The construction of such a mapping is typically carried out by learning science practitioners, such as subject matter experts, cognitive scientists and learning engineers, who inspect the materials and assign one or more KCs to each question. Once student data has been collected, the initial mapping can then be improved as poorly associated KCs come to light (Corbett and Anderson, 1994). As the next step, instructional designers are often leveraged to revise the KC model; however, this is often a time-consuming task, making continuous iteration challenging. While machine learning methodologies have been developed to assist in the automatic identification of new KCs, prior research has shown that human judgment remains critical in the interpretation of the improved model and acquisition of actionable insights (Liu and Koedinger, 2017; Nguyen et al., 2019).

An emerging area that has the potential to provide the human resources needed for scaling KC modeling is crowdsourcing. Over the last decade, crowdsourcing has become a prominent model of distributed problem-solving, especially on tasks that are doable by humans, but difficult for computers (Yuen, King, and Leung, 2011). It has also seen wide usage in academic research as a source of training data for machine learning, support for integrated workflows, and assistance to domain experts (Kobren et al., 2014). However, to our knowledge, there has been minimal prior work that employs crowdsourcing in KC model construction and refinement (Moore et al., 2020a).
Therefore, as a first step towards examining and promoting the feasibility of crowdsourced KC modeling, we studied how crowdworkers can provide the underlying KCs for an assessment activity. Using a crowdsourcing platform, we gathered participants with no background in pedagogical training and varying levels of math and English writing expertise. We then asked them to provide three KCs needed to solve a given problem in the domains of math and English writing. We took these contributions and input them to two topic models to analyze how they might be clustered in useful ways, indicative of any KCs. Our research questions are as follows:

**RQ1**: How do crowdsourced knowledge components compare to expert-generated ones in each problem?

**RQ2**: How might we leverage topic modeling to generate groupings of participant contributes indicative of expert-generated knowledge components?

From these questions, our goal is to see whether it is possible to employ crowdsourcing to generate a baseline KC model that is both interpretable and able to be translated into a more learnersourced context.

### 2. Related Work

KC models are typically developed by domain experts through Cognitive Task Analysis methods (Schraagen, Chipman, and Shalin, 2000), which lead to effective instructional designs but require substantial human efforts. In recent decades, this method has been greatly enhanced by data-driven techniques. For instance, (Corbett and Anderson, 1994) showed that identifying the “blips” or “peaks” in the KCs’ learning curves can help uncover new KCs that were not accounted for in the initial model. Other works have formalized the concept of KC mapping through q-matrix or skill matrix (Koedinger et al., 2004). Under this representation, several data mining techniques can be applied to assist in the refinement of KC models, for example by extracting relevant KCs from data and merging/splitting KCs to yield a better fit (Koedinger, McLaughlin, and Stamper, 2012). It should be noted that while these fully automated methods can potentially discover models with better performance (in terms of statistical metrics such as AIC, BIC and cross validation score), they suffer from a lack of interpretability (Stamper, Koedinger, and Mclaughlin, 2013). On the other hand, (Koedinger et al., 2013) showed that a refined KC model that results from both human judgment and computational metrics can help students reach mastery in 26% less time. More generally, as pointed out in (Liu, McLaughlin, and Koedinger, 2014), the inclusion of human factors in the KC modeling process can be advantageous, leading to actionable lessons that can be implemented in follow-up studies. Our work belongs to the same line of research, in which we try to leverage human insights, albeit from crowdworkers instead of traditional teachers and educators.

Traditionally, crowdsourcing can be carried out in either a marketplace, e.g. Mechanical Turk, a game environment, e.g. FoldIt (Cooper et al., 2010), or an online community, e.g. Stack Overflow. In this study, as a starting point, we recruit workers through the paid platform Amazon’s Mechanical Turk (AMT); ultimately, however, we would like to transition to a more community-based method. It has been shown that, by offering an online course, one can attract a crowd of thousands or more students, who can be leveraged to improve the course itself (Weld et al., 2012). This is a skilled crowd, one which has at least the basic skill sets and prerequisites for the course. The use of such crowd workers is common in educational technology, often in a way that leverages the workers’ content knowledge (Anderson, 2011; Moore et al., 2020b). Recently, crowdsourcing has become increasingly popular for content development in the education domain (Paulin and Haythornthwaite, 2016).

Companies such as Coursera and Duolingo also employ this approach to have their learners make improvements to the content as they use it. As another example, the process of crowdsourcing data from learners, or learnersourcing, has been used to identify which parts of lecture videos are confusing (Kim et al., 2014), and to describe the key instructional steps and subgoals of how-to videos (Kim, Miller, and Gajos, 2013). More recently, learnersourcing contributed to not only the annotation of existing educational content, but also the creation of novel content itself. In particular, (Williams et al., 2016) explored a crowdsourcing-based strategy towards personalized learning in which learners were asked to author paragraphs of text explaining how to solve statistics problems. The explanations generated by learners were found to be comparable in both learning benefit and rated quality to explanations produced by expert instructors.
As the fields of natural language processing and text mining continue to advance, they have seen increasingly wider usage in education technology to help automate intensive tasks (Brack et al., 2020). Previous works have looked at using different machine learning models (Pardos and Dadu, 2017; Patikorn et al., 2019) and utilizing a search engine (Karlovčec, Córdova-Sánchez, and Pardos, 2012) to tag educational content with KCs. Recent efforts have utilized topic modeling on a set of math problems from an intelligent tutoring system to assist in the labeling of KCs (Slater et al., 2017). While their initial model had promising results, there was an issue of human interpretability for the topics it produced, that may be relieved by different models (Lee et al., 2017). Much of the work in this space is focused on predicting KCs for content, after being trained on similar KC-tagged problems. Few studies have attempted to leverage text mining techniques to generate KCs for content, with no training or prediction modeling involved. A previous study made use of the same two topic modeling techniques this study utilizes, but for the context of classifying programming problems (Intisar et al., 2019). They found a series of beneficial trade-offs between the models in terms of accuracy and dimensionality, which we build upon with our work.

3. Methods

The study consists of two related experiments that differ in their domain content for a specific part of the task. The first domain is mathematics, with a focus on the area of shapes, such as squares and rectangles. The second domain is English writing, with a focus on prose style involving agents and clause topics. The math content is at the middle school level and the writing content at the undergraduate level in the United States. For both domains, we conducted an experiment using the Amazon Mechanical Turk (AMT) platform. Eighty unique crowd workers on AMT, known as turkers, completed the math experiment and sixty unique turkers completed the writing, for a total of 140 participants. Thirteen participants in the Writing experiment were removed from our analyses due to submitting invalid responses that indicated either a complete misunderstanding of the experiment instructions or behavior similar to a bot. Filtering these invalid participants left us with 47 total participants in writing, combined with the 80 from math, for a total of 127 participants. Among these participants, 55 self-identified as female and 72 as male. The mean selected age range was 35-44. All participants reported having at least a high school degree or equivalent. Additionally, none of the participants indicated that their occupation directly involved the use of geometry or writing prose. In each experiment, the tasks took roughly five minutes to complete. Participants were paid $0.75 on completion, providing a mean hourly wage of $9.

In the main task of the experiment, participants were given a word problem and asked to list three KCs that are required to solve the problem. The math experiment problem is about finding the area of a shape with two structures, while the writing experiment problem involves the revision of a sentence (Figure 1 and 2). The prompt for KCs was “As concisely as possible, please indicate a skill required to answer the above math problem about the wall.” in the math experiment and “As concisely as possible, please indicate a skill required to answer the above question that involved revising the sentence.” in the writing experiment. Note that the prompt uses the term “skill” rather than “knowledge component” to avoid jargon that may be confusing.

3.1 Math and Writing Experiments

The math word problem for which participants were asked to generate KCs is shown on the left side of Figure 1. This problem, along with the two priming questions, comes from a dataset titled ‘Geometry Area (1996-97) [KRM]’, which was used in a previous study of a geometry cognitive tutor (Stamper and Koedinger, 2011). An expert instructor familiar with the domain knowledge and KC modeling process tagged this problem with three KCs as described in Table 1. These will serve as a baseline for our comparison with the turkers’ generated KCs. The word problem for which participants were asked to generate KCs is shown on the right side of Figure 1. This problem comes from an online prose style course for freshmen and sophomores at a four-year university in the United States. Similar to the math experiment, we use four KCs provided by an expert instructor (Table 1) as a baseline for comparison
The height of a wall is 25.0’ and a 8’ x 20.0’ rectangular door is positioned on the wall such as there is 8’ of wall remaining on the left side and 4’ of the wall remaining on the right side.

Find the area of the wall to be painted. Do not paint the door. _______

Figure 1: The math and writing word problems that participants provided three skills for in both conditions in either the math or writing experiment.

Table 1. Expert-generated KCs in the math and writing experiment, domain code “M” for math and “W” for writing.

<table>
<thead>
<tr>
<th>KC (Domain)</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>compose-by-addition (M)</td>
<td>In an equation such as a + b = c, given any two of a, b or c, find the third variable.</td>
</tr>
<tr>
<td>subtract (M)</td>
<td>Subtract the area of one shape from another.</td>
</tr>
<tr>
<td>rectangle-area (M)</td>
<td>Finding the area of a rectangle shape.</td>
</tr>
<tr>
<td>id-clause (W)</td>
<td>Identify the clause-level topic of a sentence.</td>
</tr>
<tr>
<td>discourse-level (W)</td>
<td>Keep the discourse-level topic of the sentence in focus.</td>
</tr>
<tr>
<td>subject-position (W)</td>
<td>Assess whether an entity is a subject.</td>
</tr>
<tr>
<td>verb-form (W)</td>
<td>Transform a passive verb to an active verb.</td>
</tr>
</tbody>
</table>

3.2 Topic Modeling & Evaluation

Topic models estimate latent topics in a document from word occurrence frequencies, based on the assumption that certain words will appear depending on potential topics in the text. To create clusters that were easier to quickly process and identify the common themes of the crowdworker contributions, we leveraged topic modeling. The crowdsourced KC text was lemmatized and stop words were removed, using a common NLP library in Python (Bird, Klein, and Loper, 2009). No further text processing was performed on the crowdsourced KC data before running them through the topic models, as we wanted results without fine-tuning any parameters or heavily processing the data. To support replicability, the code used to generate both the topic models used in this study can be found at: https://github.com/StevenJamesMoore/ICCE2020/blob/master/topic_models.ipynb.

The first topic modeling technique we used is Latent Dirichlet Analysis (LDA (Blei, Ng, and Jordan, 2003)). LDA maps all documents, in this case the participant-generated KCs for the problems, to a set number of topics in a way such that the words in each document are captured by the topics (AlSumait et al., 2009). For this process we set the number of topics to five, as the math problem has three expert-generated KCs and the writing problem has four expert-generated KCs. We wanted a topic number similar to the actual number of expert KCs, but not limited to it, in case participants contributed potentially novel or latent KCs of the problems. Additionally, in this case the documents across our corpus are small in length, often only consisting of a few words or phrases. This is not a traditional document used by LDA, which generally has a topic characterized by a distribution over words (Yau et al., 2014). However, the model is still able to label skills from the brief text and identify relationships between the words that prove useful (Tong and Zhang, 2016).
Non-negative Matrix Factorization (NMF (Lee and Seung, 2001)) was the second topic modeling technique we used, which conceptually outputs topics made up of term clusters from the documents fed into it. NMF uses linear algebra for topic modeling by identifying the latent structure in data, the participant-generated KCs for the problems, represented as a non-negative matrix (Luo et al., 2017). Again, the topic number for this model was also set at five and the other parameters of the code can be accessed via the GitHub link. We chose this for the second topic modeling method as it successfully has been used in several related educational contexts previously, in order to map skills to assessment items and improve Q-matrices (Desmarais, 2012; Desmarais and Naceur, 2013).

The results of the topic models were then evaluated by two professional instructional designers who read the corresponding topic terms and also had access to the problem that crowdworkers provided KCs for, but they were not made aware of the expert-generated KCs that were previously used for the problems. These two experts worked collaboratively to construct interpretations of the topics, based on the topic five key terms that comprised them. The researchers were then tasked with identifying if these topic interpretation labels correspond to any of the expert-generated KCs for the problems. Therefore, the topic model outputs were interpreted by the researchers and compared to the KCs generated by experts. This was the evaluation phase of comparing the topic model outputs, ran on the crowdsourced KCs, to the ground truth expert-generated KCs for the math and writing problems.

4. Results

The five topics from both the LDA and NMF models for the math experiment data, along with the top five most common terms associated with each topic, are presented in Table 2. This table contains a topic label column, which was constructed as an interpretation of terms by the two instructional designers during the evaluation phase. The two models share similar topics and the top five terms that comprise them are also comparable. Both models include topics that involve basic arithmetic operations used in the problem, such as addition, subtraction, and multiplication. They differ in that the LDA one contains a topic mentioning the order of operations, which is applicable to the problem. The NMF model also contains two unique topics that mention calculations with double digit numbers and another that includes division. Interestingly, division is not used in the problem, while the problem does involve order of operations.

Table 2. Top 5 terms from 5 topics identified by the LDA and NMF models from the math data

<table>
<thead>
<tr>
<th>Topic #</th>
<th>Topic Terms</th>
<th>Topic Label</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LDA Model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>wall, calculation, _num, read, equation</td>
<td>Perform a calculation using the wall values</td>
</tr>
<tr>
<td>2</td>
<td>area, door, rectangle, subtract, wall</td>
<td>Subtracting the area of the door from the wall</td>
</tr>
<tr>
<td>3</td>
<td>multiplication, order, know, operation, subtract</td>
<td>Knowing to perform order of operations</td>
</tr>
<tr>
<td>4</td>
<td>number, add addition, subtract, digit</td>
<td>Adding &amp; subtracting numbers</td>
</tr>
<tr>
<td>5</td>
<td>multiply, subtraction, number, digit, double</td>
<td>Multiplying &amp; subtracting double digit numbers</td>
</tr>
<tr>
<td><strong>NMF Model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>number, multiply, add, digit, subtract</td>
<td>Multiplying &amp; subtracting double numbers</td>
</tr>
<tr>
<td>2</td>
<td>multiplication, skill, determine, double, digit</td>
<td>Multiplying double digit numbers</td>
</tr>
<tr>
<td>3</td>
<td>subtraction, skill, digit, length, width</td>
<td>Subtraction using the length and width of a shape</td>
</tr>
<tr>
<td>4</td>
<td>addition, skill, word, foot, division</td>
<td>Adding and dividing numbers</td>
</tr>
<tr>
<td>5</td>
<td>area, subtract, door, know, rectangle</td>
<td>Subtracting the door area from the wall</td>
</tr>
</tbody>
</table>
The five topics from both the LDA and NMF models for the writing experiment data, along with the top five most common terms associated with each topic, are presented in Table 2. This table also contains the topic label constructed during the evaluation phase. The topic terms between the two models are more diverse than those in the math experiment, although there is still overlap between the topics, as expected. Both models share a topic indicative of identifying the position of the sentence’s subject and agent. Relatively, they also share a topic of understanding the structure of a sentence and an understanding of how to write. Interestingly, the LDA model’s topic for understanding how to write also includes mention of the English language, which is not present in the NMF model’s corresponding topic. Another key topic difference is that the NMF model has two unique topics of identifying the passive voice of a sentence and has critical thinking and problem solving skills. The final noticeable topic difference is that the LDA model has a topic indicative of understanding verb tenses, which is not present in the NMF one.

Table 3. Top 5 terms from 5 topics identified by the LDA and NMF models from the writing data

<table>
<thead>
<tr>
<th>Topic #</th>
<th>Topic Terms</th>
<th>Topic Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDA Model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>agent, know, subject, identify, position</td>
<td>Identifying the subject and agent’s position</td>
</tr>
<tr>
<td>2</td>
<td>verb, tense, knowledge, understand, know</td>
<td>Understanding verb tenses</td>
</tr>
<tr>
<td>3</td>
<td>sentence, structure, understand, construction, determine</td>
<td>Understanding the structural components of a sentence</td>
</tr>
<tr>
<td>4</td>
<td>english, skill, understand, language, writing</td>
<td>Understanding the English language and how to write it</td>
</tr>
<tr>
<td>5</td>
<td>comment, writing, english, mean, language</td>
<td>Understanding the English language and how to write it</td>
</tr>
<tr>
<td>NMF Model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>english, comment, voice, passive, interpret</td>
<td>Identifying the passive voice of a sentence</td>
</tr>
<tr>
<td>2</td>
<td>know, agent, subject, position, identify</td>
<td>Identifying the subject and agent’s position</td>
</tr>
<tr>
<td>3</td>
<td>sentence, knowledge, structure, determine, meaning</td>
<td>Understanding the structural components of a sentence</td>
</tr>
<tr>
<td>4</td>
<td>skill, writing, paragraph, require, word</td>
<td>Knowing how to write</td>
</tr>
<tr>
<td>5</td>
<td>problem, thinking, critical, solve, math</td>
<td>Critical thinking and problem solving skills</td>
</tr>
</tbody>
</table>

After the creation of the topic interpretations during the evaluation phase, the researchers matched the topics to any corresponding expert-generated KCs for both the math and writing problems. The matching of any topics to these KCs is presented in Table 4. As depicted, each domain had expert-generated KCs that were not addressed by a topic from either model.

The math experiment’s first LDA model topic resembles the compose-by-addition KC as it mentions the calculation of values from the wall depicted in the problem. Both models mention subtracting across multiple topics, but only the second LDA topic and the fifth NFM topic address the subtract KC. To indicate this KC, the topic needed to explicitly mention subtracting the area of one shape (door) from another (wall or rectangle). The final math KC, rectangle-area, was not present in any model’s topic, although it may be argued that the previous topics for the subtract KC also include this one implicitly.

The first two expert-generated KCs for the writing problem are not identified by any topics in either model. Both contain domain-specific vocabulary, clause and discourse, which is not included in the top five terms for the topics. Topic three for both the LDA and NMF models indicates the structure of the sentence, which is a step towards these KCs, but not inclusive enough to match either of them. Both models had a single topic that matched the subject-position KC, as the topics explicitly mention the position of the subject and agent in the sentence. The final expert-generated KC, verb-form, is partially addressed by the second LDA topic, which mentions verb tense. However, the NFM model’s first topic more explicitly addresses this KC, as it specific the passive voice of the sentence.
Table 4. Expert-generated KCs in the math (M) and writing (W) experiment domains compared against the topics generated by the two models.

<table>
<thead>
<tr>
<th>KC (Domain)</th>
<th>LDA Topic #</th>
<th>NMF Topic #</th>
</tr>
</thead>
<tbody>
<tr>
<td>compose-by-addition (M)</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>subtract (M)</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>rectangle-area (M)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>id-clause (W)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>discourse-level (W)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>subject-position (W)</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>verb-form (W)</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

5. Discussion

Firstly, we wanted to better understand how we could have crowdworkers assist in the generation and mapping of KCs to problems across the math and writing domains. Using the results of the evaluation from expert-generated KC to topic, crowdworkers appear to have similar performance in terms of the number of expert KCs they were able to generate. The expert-generated KCs in the math domain had little domain vocabulary or included terms that were not mentioned in the question. Only ‘rectangle’ was explicitly used in the KCs, which is not included in the question text, but that is also a more common word and does not require domain knowledge to know. On the other hand, all four expert-generated KCs in the writing experiment contain vocabulary specific to the problem and its domain. It is not a surprise that the first two, which include clause and discourse, were not indicated by participants. These terms are jargon and not as common, particularly for those who may not have had an English writing course recently or speak it as a secondary language. Additionally, the math domain problem comes from a middle school course while the writing domain one comes from an undergraduate college course. It was surprising at first that participants were able to match the subject-position KC, however those terms are explicitly mentioned in the instructional text of the problem. It appears that the strategy some participants employed was using the instructions of the question to generate the KCs. This is akin to previous work that uses different machine learning models to predict KCs, which often employs the problem’s text as a key feature (Karlovčec et al., 2012).

Secondly, we wanted to see how topic modeling could be leveraged with crowdsourced KCs to produce results comparable to that of experts. This process yields a large amount of text data, that is both a mixture of applicable, redundant, and irrelevant all at the same time. To help make sense of this data in a less time-consuming matter, we turned to topic modeling techniques to aggregate the data. We found that the topics these models produce are indicative of several KCs for the problems that are akin to one’s experts previously generated and used for them. In the math experiment, two out of three of the expert KCs were produced between both models. In the writing experiment, another two of the expert KCs were produced out of four in total. Ultimately both models across each domain were comparable, resulting in almost identical topics that related back to an expert-generated KC. The LDA model for math had a slight gain over the NMF one, in that it had a topic comparable to another KC that the NMF one did not. However, this was perhaps the weakest topic-to-KC matching, so we cannot claim LDA performs better than NMF or vice-versa for this task.

The difference in topics between the two models for the math domain were not useful in the identification of an expert-generated KC. These different topics between the two addressed aspects that were still relevant to the problem, such as double-digit numbers and order of operations, but they were not the specified KCs for the problem. This was also true for the topics both math models shared, while they were mostly applicable to the problem due to indicating multiplication and addition, they were not relevant KCs. The LDA and NFM models in the writing domain had similar results, in that the topics that differed between the two pertained to the problem, but were not a KC for it. Interestingly, the LDA model for the writing domain was the only model that had a duplicate topic within its own top five topics, which was about understanding and writing the English language. While that may suggest participants thought such a skill was applicable to the problem, and it is, it did not match an expert-generated KC. While that is a relevant skill needed for the problem, it is not a KC the problem intends to measure, as the target learner is that in an undergraduate writing course.
5.1 Limitations

In our study, we utilized two forms of topic modeling, LDA and NMF, to group potential KCs contributed by crowdworkers. Different techniques for topic modeling may produce more accurate results or ones that are more interpretable by humans (Chang et al., 2009). While performing this modeling on the data helps to aggregate it into a more manageable fashion, human interpretation is still heavily required to make use of the data. Additionally, to keep the data collection brief, participants did not receive explicit instruction on the desired granularity or process of generating KCs. Measuring a participant’s existing domain knowledge and training them on this task could produce more reliable results (Koedinger and Nathan, 2004). Finally, while we had multiple researchers and experts involved in the process of interpreting the topics and evaluating their matches to expert-generated KCs, there is still a subjective element at play with what topic or KC a contribution may fall under. Making use of existing text mining techniques for this may help to increase replicability of this process (Wiedemann, 2016), along with including more data coders in this process.

6. Conclusion and Future Work

In this work, we solicited the knowledge components of a mathematics problem and an English writing problem from crowdworkers with different expertise and backgrounds. Our results indicate that crowdworkers can be effectively leveraged to assist in the process of generating knowledge components for problems in our math and writing domains. To our knowledge, this work is among the first to investigate the use of crowd workers to generate knowledge components for assessment activities. We found that roughly half of the KCs generated by experts and previously used for the problems in an educational setting were able to be matched by KCs generated by the crowdworkers. The LDA and NMF topic models created groupings that when interpreted by a human, were akin to the expert generated ones. Several of these topics proved useful, as they indicated KCs of the problems, but work remains to improve both interpretability and matching more domain-specific KCs. Ultimately the results of this work support the benefits of engaging the crowd in more nuanced learning science tasks, as well as encourage the adoption of a similar process that utilizes learnersourcing. With continued improvement of the efficiency and efficacy of this process, we envision further use of crowdworkers to generate high quality and accurate knowledge components that easily scale up to benefit both learners and education providers.

Building upon this work, we plan to integrate this process in a learnersourced context, where participants (i.e., students) potentially have more commitment and domain knowledge that could be leveraged (Paulin and Haythornthwaite, 2016). This would enable us to properly train them to provide such KCs throughout the course, rather than completing the task once with only a brief instruction like the crowdworkers did in this study. Ultimately, we envision a workflow in which students submit KCs for problems; these problems are then peer reviewed and presented to the teachers (or relevant parties) to help them confirm applicable KCs and improve the assessment items. This procedure is analogous to the find-fix-verify pattern in crowdsourcing, which has been shown to be effective (Bernstein, 2013). This study demonstrates the first step in developing such a workflow, providing initial insights into how crowdsourced contributions might be leveraged in the KC mapping process. While the analyses we performed here may end up taking more time than the task of generating the KCs themselves, they contribute to an eventual domain-independent workflow that can greatly expedite such tasks.

Our findings also suggest several directions for future research in crowdsourcing KCs and involving crowdworkers in instructional design tasks without the need for prior training. Performing a similar study involving different domains at various grade levels might yield interesting results as well, especially when crowdworkers might have more familiarity thanks to their diverse backgrounds. Additionally, studying the impact of grouping crowdworkers on their reported expertise levels and then having them generate KCs for problems of varying difficulty levels could introduce useful design decisions for implementing this workflow into a digital learning environment. Ultimately, we would like to scale this up into a learnersourced context, such as embedding these prompts into an online course. Our next step is to leverage different natural language processing techniques outside of topic modeling, such as keyphrase extraction or summarization, in order to achieve higher accuracy and reduce the need for human input into the interpretability aspect of this work.
References


Primary students’ readiness for learning of artificial intelligence: A case study in Beijing

Ching Sing CHAI<sup>a</sup>, Pei-Yi LIN<sup>a</sup>, Morris Siu-Yung JONG<sup>ab</sup>, Yun DAI<sup>a</sup> & Thomas CHIU<sup>ab</sup>

<sup>a</sup>Department of Curriculum and Instruction & <sup>b</sup>Centre for Learning Sciences and Technologies
The Chinese University of Hong Kong, Hong Kong

Abstract: Recent advances in artificial intelligence (AI) has aroused attention among educators to consider including it in K-12 teaching. Several curricula have been developed, and some schools have begun testing the curriculum. Within such a context, it is important to study and investigate students’ readiness to learn AI. This study includes students’ motivation to learn AI, perceived relevance of AI, anxiety about current or future development of AI to predict students’ AI readiness. A questionnaire-based survey composed of these factors was designed. A sample of 107 primary students from Beijing, China was used to provide initial evidence about the validity and reliability of the questionnaire. The findings support that the questionnaire has adequate construct validity and acceptable reliability. In addition, the findings indicate that students have intrinsic motivation and career motivation. They also perceive the relevance of AI and regard themselves are ready for an AI-empowered future.

Keywords: artificial intelligence, readiness, motivation, relevance, anxiety

1. Introduction

Artificial intelligence (AI) has been viewed as the next effective tools to enhance teaching and learning. AI educational research has been focused on AI development and applications (for a review, see Roll & Wylie, 2016). For instance, an intelligent tutoring system (Ma, Adesope, Nesbit, & Liu, 2014), an adaptive learning system (Nakic, Granic, & Glavinic, 2015), and such AI technologies are used in K-12 and university education settings to support learning. Beyond education, AI is an emerging technology and AI applications have pervasive influences in all areas. AI is also developed as a new curriculum to be learned in K-12 settings. Given the background, a fundamental question arises: how to promote student’s readiness to learn AI? Current AI learning research has not sufficiently explored primary students learning about AI.

Technology readiness reflects an individual’s willingness to leverage new technologies to accomplish tasks (Parasuraman, 2000). Investigations of students’ technology readiness reflect how curricula are preparing students to live and thrive with the technology. This would usually entail studying how psychological factors associated with curriculum design contribute to the students’ sense of readiness. Strong curriculum design needs to enhance students’ motivation to learn (Yilmaz, 2017) and establish relevance of learning for students (Jong, 2014, 2020). These conditions, when met, help students to learn about specific technology and thus build their readiness. In the case of AI, strong curriculum design would also need to reduce students’ fear of AI (Wang & Wang, 2019) and promote a possibility for students to choose AI engineering as a future career choice. Surveying these variables that influence students’ readiness allows educators to refine the teaching and learning strategies used to promote students’ readiness to learn AI. This study considered the potential effects of intrinsic motivation, career motivation, perception of relevance of AI, and AI anxiety on students’ readiness in the AI learning context when primary students were introduced to AI curriculum.

2. Literature review
2.1. AI readiness

AI readiness refers to the construct of technology readiness (Parasuraman, 2000) to unpack students’ propensity to learn AI. Meanwhile, readiness is an important variable in indicating student learning mediated with or by technologies, if students are more techno-ready, they will tend to engage in technology learning and seek technology-related professions (Parasuraman & Colby, 2015). Past research indicates students technology readiness is an essential variable to ascertain students preference to learn new technologies (e.g., e-learning, see Gay, 2016; mobile learning, see Cheon et al., 2012; game-based learning, see Jong et al., 2006). AI has widened the scope of learning in education by adopting AI technology. AI readiness construct can be viewed as an overall state of mental processes resulting from how students’ predisposition to learn AI as learning content (Lin, Shih, & Sher, 2007). This study refers to technology readiness index confirmed by Parasuraman (2000), and adapts items to measure students’ readiness to learn AI.

2.2. Motivation for AI learning

Learning motivation is an important factor to explain the process that people invest in learning activities and to explain their willingness to engage in a learning environment (Jong et al., 2010; Shang et al., 2006). Students’ readiness is identified by types of motivations, attitudes, and behaviors in a computer-supported collaborative learning environment (Xiong, So, & Toh, 2015). In other words, motivations for AI learning could be viewed as a component of measuring students’ readiness to engage in AI learning. Ryan and Deci (2000) indicated that intrinsic motivation is an element attached to learning processes. An individual is intrinsically motivated when (s)he feels interesting, gains pleasure and satisfaction from completing or working on a challenge. Extrinsic motivation is related to an individual’s desire to succeed and achieve certain goals (Ryan & Deci, 2000). That is, increased student motivations will enhance willingness to learn AI technologies. In the present study, we measure students’ intrinsic motivation to learn AI out of curiosity and interest. Extrinsically career motivation associated with AI learning, on the other hand, should include students’ career decisions and behaviors, such as searching for and accepting a job. Past motivation study has shown that students’ level of e-learning readiness and motivation are associated (Yilmaz, 2017). In this study, both intrinsic and extrinsic motivations are assumed to predict AI readiness.

2.3. Relevance of AI

Relevance refers to that a person perceives learning content meets his/her needs or desires (Keller, 2000). Students who perceive AI learning content is personally relevant or is connected to their goals will be motivated to participate in AI learning. That is, if students feel the AI learning content is meaningful, or recognize the importance of AI will help them achieve personal goals, motives, and values, then they will be motivated to learn AI. Relevance can also be understood pedagogically as a factor that illustrates people’s actual experiences when they can relate prior knowledge and experiences to the current content they learned, and when the content can be applied on their jobs or in real life (Keller, 2010). Park, Nam, and Cha’ study (2012) pointed out that when university students found that mobile learning was relevant to their study and future career, they tended to use mobile to learn. In the context of AI learning, relevance of AI refers to students’ ability to connect the AI learning content as something useful or meaningful to them. When students see the connections between what they need to know about AI and what AI learning will bring to them, they will learn more about AI, and this will thus build their readiness for an AI-empowered world.

2.4. AI anxiety

AI anxiety refers to an effective response of feeling of fear being expressed about out-of-control AI (Johnson & Verdicchio, 2017). Past research indicated that people’s fear and insecurity towards technologies influence people’s adoption of technologies. This study includes AI anxiety to understand the impact of advanced AI technology on students’ readiness for the AI-empowered future. As AI technology will certainly transform how we live and work, people will be required to
learn in-demand knowledge and update their skills in order to remain relevant and meet future career needs (Korinek & Stiglitz, 2017). Meanwhile, they may feel anxious toward AI when they perceive that AI technology will change or eliminate jobs and create new ones (Wang & Wang, 2019). Previous research on ICT anxiety was found to have a relation to students and educators’ use of mobile learning (Mac Callum, Jeffrey, & Kathryn, 2014). In addition, technology may cause students to feel anxious because they had negative or inadequate experience in learning to use the technology (Keller, 2010). Similarly, anxiety toward AI technology may influence its uses. As a result, this study assumed that self-perceived fear about AI technology would have a negative influence on people’s sense of readiness to learn AI.

2.5 Gender Differences

Lee, Chai and Hong (2019) review of existing STEM research indicates that gender differences are an important issue for STEM-related disciplines, which is echoed in some recent related literature (e.g., So et al., 2020). AI, as a discipline, is situated within the field of engineering and the design of AI applications depends much on the engineers’ knowledge about science and mathematics. It is likely that there may be gender differences between students when they learn about AI. Hence, this study also investigates gender differences among the students.

In sum, this study examines the following two research questions: (1) Is the 5-factor AI learning model reliable and valid? (2) Do the variables, i.e., intrinsic and extrinsic motivations, relevance of AI and AI anxiety predict primary students AI readiness? (3) Are there gender differences among the measured variables?

3. Method

3.1. Participants

The current sample included 107 students who were participated in AI learning in Beijing, China. They were third to sixth graders from two primary schools. The sample included 57 females and 50 males with an overall mean age of 9.08 (SD = 0.72). The mean hours spent on AI-learning and projects were 6.75 (SD = 3.09). The AI curriculum covered AI developments and applications such as the history of AI, image and voice recognition, content recommendation, machine learning, and ethical issues. At the end of the semester, the participants were informed about the purpose of the anonymous survey, and participation was voluntary. They were instructed to respond to each item by choosing the response that accurately described their level of agreement.

3.2. Instruments

This study included five subscales: (1) intrinsic motivation subscale was adapted from motivated strategies for learning questionnaire (MSLQ) – intrinsic goal orientation (Pintrich, Smith, Garcia, & McKeachie, 1991) (4 items; e. g., “In the AI class, I prefer course material that arouses my curiosity, even if it is difficult to learn.”). (2) Career motivation subscale was adapted from Glynn, Taasoobshirazi, and Brickman (2009) (3 items; e.g., “I think about how learning AI can help me get a good job.”). (3) Relevance of AI subscale were selected and modified from Keller’s (2010) the Instructional Materials Motivation Survey and the Course Interest Survey (6 items; e. g., “The things I am learning in this AI class will be useful to me.”). (4) AI Anxiety subscale was modified from MSLQ – test anxiety (Pintrich et al., 1991) (5 items; e. g., “When I consider the capabilities of AI, I think about how difficult my future will be.”). (5) AI readiness subscale was selected and modified from Parasuraman’s (2000) technology readiness index, we adapted 6 items in the present study (e. g., AI products and services that use the newest technologies are much more convenient to use.). The items in the questionnaires were presented on a 4-point Likert scale (1 = strongly disagree; 4 = strongly agree). After experts in AI-learning reviewed the first draft of these items, school teachers were invited to review the questionnaires and provided their feedbacks. The final versions of the questionnaire were then confirmed.
3.3. Data analysis

Exploratory factor analysis (EFA) was employed to examine the 5-factor AI learning model. The principal axis factoring was utilized as the extraction method, along with the rotation method of direct oblimin. The alpha reliabilities of the scales were generated to check reliability. Correlation and regression analysis were then conducted to examine whether predictors: intrinsic motivation, career motivation, relevance of AI, and AI anxiety factors could predict AI readiness (outcome variable).

4. Results

4.1. Exploratory Factor Analysis of the Measurement Model

As shown in Table 1, a total of 18 items were retained in the final version of the measurement. Six items with cross-loadings were omitted and five factors were revealed: ‘intrinsic motivation’ (IM), ‘career motivation’ (CM), ‘relevance of AI’ (R), ‘AI anxiety’ (A), and ‘AI readiness’ (RE). Items factor loading ranged from 0.46 to 0.93. The total variance explained was 70.47%. Cronbach’s alpha coefficient of these factors was around 0.81 – 0.95 for each scale, and the overall reliability was 0.80, which revealed high reliability of these factors, suggesting that the internal consistency was sufficient for statistical analysis. Table 1 presents the results of descriptive statistics and EFA.

Table 1. EFA of measured items

<table>
<thead>
<tr>
<th>Factor</th>
<th>M</th>
<th>SD</th>
<th>α</th>
<th>Factor loading</th>
<th>% of variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Intrinsic motivation (IM)</td>
<td>3.51</td>
<td>0.64</td>
<td>0.85</td>
<td>0.46 - 0.65</td>
<td>3.20</td>
</tr>
<tr>
<td>2. Career motivation (CM)</td>
<td>3.46</td>
<td>0.63</td>
<td>0.88</td>
<td>0.67 - 0.82</td>
<td>38.99</td>
</tr>
<tr>
<td>3. Relevance (R)</td>
<td>3.55</td>
<td>0.57</td>
<td>0.81</td>
<td>0.47 - 0.84</td>
<td>4.44</td>
</tr>
<tr>
<td>4. Anxiety (A)</td>
<td>2.25</td>
<td>1.09</td>
<td>0.95</td>
<td>0.79 - 0.93</td>
<td>18.78</td>
</tr>
<tr>
<td>5. Readiness (RE)</td>
<td>3.62</td>
<td>0.49</td>
<td>0.83</td>
<td>0.56 - 0.92</td>
<td>5.06</td>
</tr>
</tbody>
</table>

4.2. Correlation Analysis among the Subscales

The Pearson correlation coefficients between the factors were calculated to explore the relationship between students’ perceptions of AI learning motivations, relevance of AI, AI anxiety and AI readiness. The intrinsic motivation, career motivation, relevance of AI, and AI readiness were significantly and positively related (from $r = 0.55$ to $r = 0.65$, $p < .001$). AI anxiety was significantly and negatively related to intrinsic motivation ($r = -0.21$, $p < .05$), relevance of AI ($r = -0.26$, $p < .001$) and AI readiness ($r = -0.20$, $p < .05$). The results of correlation analysis provided initial support for the relationships between the measured factors.

4.3. Predicting students’ readiness to learn AI

The linear regression analysis was conducted to predict students’ readiness to learn AI. The results showed that intrinsic motivation was the strongest predictor for all the factors ($\beta = 0.32$, $p < 0.01$), relevance of AI and career motivation are equally contributing to AI readiness ($\beta = 0.23$, $p < 0.05$). The findings indicate that the students’ motivations for learning AI and their perception of the relevance of AI provided 43.8% explanation, and these variables are very important predictors for students’ readiness for AI learning and technology.
4.4. Gender differences among measured variables

Independent t-tests were employed to compare male and female students’ perceptions of AI readiness, motivation for AI learning, relevance of AI, and AI anxiety. Male students significantly possessed greater career motivation ($t = 2.66, p < .01$) and readiness ($t = 3.30, p < .01$) for AI than female students. There were no significant gender differences among intrinsic motivation ($t = 1.48, p > 0.5$), relevance of AI ($t = 1.81, p > 0.5$) and AI anxiety ($t = 1.05, p > 0.5$). This suggests that male students were more likely to be ready for AI, and they have greater career motivation to engage in AI professions than female students.

5. Discussion

The rapid growth of AI technology will push educational institutions to adopt relevant AI curriculum. There is a need to understand what factors will influence students’ readiness for AI learning. Based on our survey, we validated a questionnaire with the five factors including intrinsic and career motivations for AI learning, perceptions of relevance of AI, AI anxiety and AI readiness, and explored the relationships among these factors. This study provides the following findings. First, the survey indicates that the questionnaire about students’ AI learning with five factors demonstrated satisfactory reliability with construct validity. This indicated that it is an acceptable instrument to assess students’ readiness to learn AI. This study incorporated five factors, and they were successfully maintained through EFA.

Second, according to the initial research findings, the analyses highlight that intrinsic and career motivations, relevance of AI predict students’ readiness. Meanwhile, AI anxiety did not significantly predict AI readiness. The results imply that students who are more intrinsically and extrinsically motivated to learn AI, and who perceived AI as relevant to them, are readier for an AI-empowered world. Previous research indicated that motivation is an important factor for teachers’ readiness to integrate technology in teaching and learning (Copriady, 2014), and students’ motivation toward e-learning is one of the component of e-learning readiness (Yilmaz, 2017). This study provides evidence that motivations are important factors in predicting students’ AI readiness. Few studies attempted to adopt relevance of technology as a predictor in a technology-based learning environment (Chai et al., in press; Dai et al., 2020; Park et al., 2012). This study considered relevance of AI is an essential factor in the AI learning context, and the finding showed that relevance of AI significantly predicted students’ AI readiness. Our findings may offer AI course designers and instructors a means for examining their courses in terms of motivational design, relevance and corresponding effects on enhancing students’ readiness. They may also need to pay attention to possible gender differences.

Acknowledgements

We appreciate the support of the school principals given to the study, as well as the participation of the teachers and students in the experiment.

References


Predicting Student Success for Programming Courses in a Fully Online Learning Environment

Neil Arvin BRETTANA*, Mehdi ROBATIb, Aastha RAWATb, Aashi PANDEYb, Shreya KHATRIb, Kritika KAUSHALb, Sidarth NAIRb, Gerald CHEANGb, Rhoda ABADIAa

aUniSA Online, University of South Australia, Australia
bUniSA STEM, University of South Australia, Australia
*neil.bretana@unisa.edu.au

Abstract: The emergence of online learning environments is important for teaching programming courses. In this study, demographic and performance-related data from two programming courses of a fully online learning platform, UniSA Online, were explored. Statistically significant features were identified using Variance Inflation Factor and Chi-Square test. Four prediction models were trained and tested using four sets of features: demographic, performance, statistically significant features, and all available features. The model trained using demographic features yielded an accuracy of 45.45%. The models trained using performance-related features, statistically significant features, and all features yielded an accuracy of 86.86%, 86.53%, and 86.53%, respectively. This highlights the importance of performance-related data in predicting student success outcomes in learning programming via a fully online learning environment.

Keywords: Online learning, model, classifier, student success

1. Introduction

Online learning has become a viable alternative for learners who are unable to participate in a traditional face-to-face university (Zhou, 2016). Online learning environments are presumed to be more inclusive relative to traditional learning environments as it allows participants of all ages, education levels, diverse regional backgrounds and even those whose performance may be limited due to accessibility needs. It is important to improve student success and learning experience and increase student retention to help achieve the goals of both the students and the learning providers. In order to achieve these, learning providers must understand their students, be able to intervene challenges early on, engage effectively with the students, and support the students through content and delivery (Stone, 2017). Aslanian and Clinefelter (2013) have shown that the educational outcomes in online learning environments are linked to the reputation of both facilitators and institutions.

Given the increase in the popularity of online learning platforms, it has become crucial to understand what characteristics affect student success in online learning environments to help guide facilitators. The growing amount of educational data in online learning provides challenges for online facilitators to organize and understand these large complex data set. With limited resources available, this large amount of data is making it increasingly difficult to monitor, identify, and solve learning issues ahead of time (Howarth, 2019; Koutropoulos & Zaharias, 2015).

Using predictive analytics in online learning can assist in predicting student performance that can assist in academic retention and student support efforts, especially for at-risk students. The use of predictive analytics approach can help understand what characteristics affect student success in online learning environments. Previous studies have shown a strong association between online learners’ performance and their demographic characteristics, such as regional belonging, socio-economic standing, education level, gender and age (Rizvi, Rienties, & Khoja, 2019). Many studies have also observed learners’ behavioural changes in online learning environments over time (Kloft, Stiehler, Zheng, & Pinkwart, 2014; Nguyen,
Huptych, & Rienties, 2018). However, these mainly relied on clickstream information from the respective week using data obtained from Massive Open Online Courses (MOOC). Although clickstream information can provide information about how a student navigates through and interacts with online education resources, studies that use clickstream data focus only on understanding student's self-regulatory behaviours, i.e., how students are using the online resources to improve instructional design (Bodily & Verbert, 2017; Diana et al., 2017; Paechter & Maier, 2010; C. Shi, Fu, Chen, & Qu, 2015) and identifying "stop-out classifiers "to prevent students from quitting (Whitehill, Williams, Lopez, Coleman, & Reich, 2015). There were also studies on the early detection systems for poor course performance (Baker, Lindrum, Lindrum, & Perkowski, 2015) but these predictive models are based on indicators that are meaningful only to academic staff who are teaching history courses and may not be useful for other courses. For instance, these wouldn't be applicable to analyzing online programming courses, which are major courses in many online learning environments. Where other courses look at high activity in discussion forums as formative assessments, programming courses give more value to practical coding formative assessment than postings in discussion forums (Azmi, Ahmad, Iahad, & Yusof, 2017; Restrepo-Calle, Ramírez Echeverry, & González, 2019).

Several studies have looked at predicting student success on a programming course using online data, but these studies were focused on a hybrid course delivery where students were required to be in the university and use an online learning environment to learn the course (Azcona, Hsiao, & Smeaton, 2019; Azcona & Smeaton, 2017; Carter, Hundhausen, & Adesope, 2017). Azcona et al. (2019) detected student at-risk by looking at students' demographics (age, travel distance from home to the university and basis of admission) and digital behaviour log. Azcona and Smeaton (2017) looked at student engagement and effort as predictors. And Carter et al. (2017) explored the relationship between the students' programming behaviours and course outcomes, and students' participation with the online social learning environment and course outcomes. While Yukselturk and Bulut (2007) studied students in fully online course, their analysis focused on identifying predictors of student's success and no predictive model was created to identify which students will most likely succeed or are at-risk of failure in the course.

In this paper, data from two fully online programming courses were considered. Data were collected from UniSA Online – a fully online learning platform operated by the University of South Australia. A model for predicting student success was trained and tested with the main aim of informing which features indicate likely success in a fully online programming course. The results of this work aim to further enable online academic staff in programming courses to identify and support students most at-risk of failing.

2. Methodology

2.1 Cohort

University of South Australia (UniSA) is one of Australia's largest online education provider. As part of its digital learning strategy, it launched the UniSA Online to provide 100% online degrees. One of the main online degrees offered through UniSA Online is the Bachelor of IT and Data Analytics. In this degree program, students study a series of programming courses in a fully online learning environment. Since the courses are fully online, learning is asynchronous. All course contents are made available to the students from the first day of the study period. Throughout the course, UniSA Online students get focused and personalized support from dedicated online academic staff via various forms of communication and quality assessment feedback presented in various forms. Students interact with their peers and the academic staff through course forums, live chats, regularly scheduled video conferencing sessions and course e-mails.

Two introductory programming courses in the online degree were chosen for this study: Python programming (Course ID: COMP 1043) and Java and Object-Oriented Programming (Course ID: COMP 1044). For COMP 1043, the final grade was calculated from the following: Programming Assignment 1 (10%), Programming Assignment 2 (15%), Programming Assignment 3A (15%), Programming Assignment 3B (15%), Programming Assignment 3C (15%), Programming Assignment 4 (15%), Programming Assignment 5 (15%), Programming Assignment 6 (15%), Programming Assignment 7 (15%), Programming Assignment 8 (15%), Programming Assignment 9 (15%).
Assignment 3B (10%), and a final exam (50%). For COMP 1044, the final grade was calculated from the following: Continuous Assessment (10%), Assignment A (25%), Assignment B (25%), and a final exam (40%). For both courses, success is defined by having 45% and above score in the online exam and having a total final grade of greater than or equal to 50.

Although the programming language taught in these courses are different, the format of the online courses is highly similar. Content videos and code-along videos are presented to the students in addition to the e-readings. Students have regular online practical activities and regularly scheduled video conferencing sessions which are recorded and available for students to watch.

2.2 Data pre-processing

Data were collected from UniSA online students who completed COMP 1043 and COMP 1044 in consecutive study periods in 2018 and 2019. From a set of 341 student records, 33 student records showing withdrawal from the course have been removed, resulting in 308 students for the final data set. Combining all remaining records for the two courses, the final data set consisted of 297 records. As this study aims to look at the difference between passing and failing the course, the final grade for each student were recorded as a binary value. To be able to feed the information into the model, the features selected for this study were also converted into binary values.

A set of 24 features and 1 identifier were engineered for each student (Table 1). The set of features were categorized into demographics and performance-related attributes pertaining to quizzes, assignments, exam and programming exercises. It should be noted that the age bins followed the standard used by UniSA online based on student professional experience. The feature Failed, which represents whether or not a student failed to succeed in the course, was set as the target feature for the model. For this study, students who had to do a supplementary assessment at the end of the study period to increase their final grade to 50 were still assigned a Failed feature value of 1 to represent a grade below 50. This is because supplementary assessments are given to students on a case-by-case basis to ultimately pass the course if they have originally obtained a failing mark.

2.3 Identifying statistically significant features

A series of statistical tests were employed to identify statistically significant features associated with student success in this setting.

2.3.1 Variance Inflation Factor

Variance Inflation Factor (VIF) is a measure applied to check multi-collinearity between available features. Multi-collinearity, represented by a high variance, makes it difficult to differentiate between the effects of supposedly independent features on the target variable (Jayaprakash, Moody, Lauria, Regan, & Baron, 2014). It also gives rise to feature redundancy, which means that a feature shares the same linear dependency as other features and does not contribute to the improvement of model performance (Brooks, Thompson, & Teasley, 2015). Therefore, it’s important to remove highly correlated features in order to enhance model performance and avoid feature redundancy. VIF achieves this by assigning a score to each feature as defined by the following equation:

\[ VIF = \frac{1}{1 - R^2} \]
<table>
<thead>
<tr>
<th>Feature</th>
<th>Type of Feature</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student_ID</td>
<td>Identifier</td>
<td>9 integers</td>
<td>A unique identifier for each student.</td>
</tr>
<tr>
<td>Gender_Male</td>
<td>Demographic</td>
<td>0 or 1</td>
<td>A binary value representing student gender: 0 (female), and 1 (male)</td>
</tr>
<tr>
<td>Age_Binned_1</td>
<td>Demographic</td>
<td>0 or 1</td>
<td>A binary value representing whether student age is 21 or under: 0 (no), and 1 (yes)</td>
</tr>
<tr>
<td>Age_Binned_2</td>
<td>Demographic</td>
<td>0 or 1</td>
<td>A binary value representing whether student age is 22 to 24: 0 (no), and 1 (yes)</td>
</tr>
<tr>
<td>Age_Binned_3</td>
<td>Demographic</td>
<td>0 or 1</td>
<td>A binary value representing whether student age is 25 to 29: 0 (no), and 1 (yes)</td>
</tr>
<tr>
<td>Age_Binned_4</td>
<td>Demographic</td>
<td>0 or 1</td>
<td>A binary value representing whether student age is 30 to 39: 0 (no), and 1 (yes)</td>
</tr>
<tr>
<td>Age_Binned_5</td>
<td>Demographic</td>
<td>0 or 1</td>
<td>A binary value representing whether student age is 40 to 49: 0 (no), and 1 (yes)</td>
</tr>
<tr>
<td>Age_Binned_6</td>
<td>Demographic</td>
<td>0 or 1</td>
<td>A binary value representing whether student age is 50 or above: 0 (no), and 1 (yes)</td>
</tr>
<tr>
<td>Location_NSWS</td>
<td>Demographic</td>
<td>0 or 1</td>
<td>A binary value representing whether student location is in the state of New South Wales: 0 (no), and 1 (yes)</td>
</tr>
<tr>
<td>Location_SA</td>
<td>Demographic</td>
<td>0 or 1</td>
<td>A binary value representing whether student location is in the state of South Australia: 0 (no), and 1 (yes)</td>
</tr>
<tr>
<td>Location_NT</td>
<td>Demographic</td>
<td>0 or 1</td>
<td>A binary value representing whether student location is in the state of Northern Territory: 0 (no), and 1 (yes)</td>
</tr>
<tr>
<td>Location_QLD</td>
<td>Demographic</td>
<td>0 or 1</td>
<td>A binary value representing whether student location is in the state of Queensland: 0 (no), and 1 (yes)</td>
</tr>
<tr>
<td>Location_TAS</td>
<td>Demographic</td>
<td>0 or 1</td>
<td>A binary value representing whether student location is in the state of Tasmania: 0 (no), and 1 (yes)</td>
</tr>
<tr>
<td>Location_VIC</td>
<td>Demographic</td>
<td>0 or 1</td>
<td>A binary value representing whether student location is in the state of Victoria: 0 (no), and 1 (yes)</td>
</tr>
<tr>
<td>Location_WA</td>
<td>Demographic</td>
<td>0 or 1</td>
<td>A binary value representing whether student location is in the state of Western Australia: 0 (no), and 1 (yes)</td>
</tr>
<tr>
<td>Location_Overseas</td>
<td>Demographic</td>
<td>0 or 1</td>
<td>A binary value representing whether student location is outside of Australia: 0 (no), and 1 (yes)</td>
</tr>
<tr>
<td>Part_Time</td>
<td>Demographic</td>
<td>0 or 1</td>
<td>A binary value representing a mode of study: 0 (no), and 1 (yes)</td>
</tr>
<tr>
<td>Failed</td>
<td>Performance</td>
<td>0 or 1</td>
<td>A binary value representing student failure as an outcome of the course: 0 (no), and 1 (yes)</td>
</tr>
<tr>
<td>-----------------------</td>
<td>----------------</td>
<td>--------</td>
<td>-------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>All_Assessment_Submitted</td>
<td>Performance</td>
<td>0 or 1</td>
<td>A binary value representing whether a student submitted all marked assessments for a course: 0 (no), and 1 (yes)</td>
</tr>
<tr>
<td>No_Assessment_Submitted</td>
<td>Performance</td>
<td>0 or 1</td>
<td>A binary value representing whether a student has submitted none of the marked assessment for a course: 0 (no), and 1 (yes)</td>
</tr>
<tr>
<td>All_Mandatory_Quizzes_Submitted</td>
<td>Performance</td>
<td>0 or 1</td>
<td>A binary value representing whether a student has submitted all mandatory quizzes for a course: 0 (no), and 1 (yes)</td>
</tr>
<tr>
<td>All_Mandatory_Assignments_Submitted</td>
<td>Performance</td>
<td>0</td>
<td>A binary value representing whether a student has submitted all mandatory assignments for a course: 0 (no), and 1 (yes)</td>
</tr>
<tr>
<td>Completed_Exam</td>
<td>Performance</td>
<td>0</td>
<td>A binary value representing whether a student has completed the exam for a course: 0 (no), and 1 (yes)</td>
</tr>
<tr>
<td>Activity_Count_0</td>
<td>Performance</td>
<td>0</td>
<td>A binary value representing whether a student failed to complete any non-mandatory activity for the course: 0 (no), and 1 (yes)</td>
</tr>
<tr>
<td>Activity_Count_1</td>
<td>Performance</td>
<td>0/1</td>
<td>A binary value representing whether a student was able to complete at least 1 non-mandatory activity for the course: 0 (no), and 1 (yes)</td>
</tr>
</tbody>
</table>

**Table 2. Summary of final grades per course and study period**

<table>
<thead>
<tr>
<th>Course ID</th>
<th>Study period</th>
<th>Pass</th>
<th>Fail</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMP 1043</td>
<td>1</td>
<td>42</td>
<td>23</td>
<td>65</td>
</tr>
<tr>
<td>COMP 1043</td>
<td>2</td>
<td>68</td>
<td>40</td>
<td>108</td>
</tr>
<tr>
<td>COMP 1044</td>
<td>1</td>
<td>36</td>
<td>5</td>
<td>41</td>
</tr>
<tr>
<td>COMP 1044</td>
<td>2</td>
<td>11</td>
<td>19</td>
<td>30</td>
</tr>
<tr>
<td>COMP 1044</td>
<td>3</td>
<td>27</td>
<td>26</td>
<td>53</td>
</tr>
</tbody>
</table>
For each feature, linear regression is performed against the other remaining features to get the value of $R^2$ (Kutner, Nachtsheim, Neter, & Li, 2005). Using this value, the VIF score is calculated using the equation above. The higher the value, the more correlated the feature is to the other features. A threshold of VIF greater than or equal to 10 has been selected to reject features (Alauddin & Nghiem, 2010; Midi & Bagheri, 2010). Features with a VIF score meeting this threshold were excluded. A recursive method was performed with the removal of features over the threshold at every step until all factors are given a score below the threshold.

### 2.3.2 Chi-square

Using the remaining features selected via the VIF test, a chi-square test was conducted to check for association with the target variable Failed. The null hypothesis of this test is that no relationship exists between the variables. However, if the p-value hypothesis of this test is that no relationship exists between the variables. However, if the p-value is higher than the defined alpha value, then there is enough evidence to reject the null hypothesis and state that there is a relationship between two values (Novaković, 2016).

### 2.4 Model training and testing

#### 2.4.1 Support Vector Machine

Support Vector Machine (SVM) was employed to build the models for this study. SVM maps nonlinear data to a higher-dimensional linear space where data can be linearly classified by hyperplane (Z. Shi, 2011). For this study, the SVM model implementation in RapidMiner (Mierswa & Klinkenberg, 2018) was utilized with a linear Kernel type and a cost of misclassification value of 0.01.

#### 2.4.2 Training

Four training sets were prepared based on the type of features. The first set included only demographic features, the second set included only performance features, the third set included only statistically significant features resulting from the mutual info gain test, and the last set included all features. Four models were trained based on these sets.

#### 2.4.3 Testing

To test the predictive performance of the constructed models, a 10-fold cross-validation was carried out. Each training set was divided into 10 groups by splitting each data set into 10 approximately equal-sized subgroups using stratified sampling. During cross-validation, each one of the 10 subgroups was regarded as the validation set in turn, and the remainder was regarded as the training set. The average of each run was calculated.

The following measures of predictive performance of the trained models were calculated: Precision (Pre) = TP/(TP+FP), Sensitivity (Sn) = TP/(TP+FN), Specificity (Sp) = TN/(TN+FP), and Accuracy (Acc) = (TP + TN)/(TP+FP+TN+FN), where TP, TN, FP and FN represent the numbers of true positives, true negatives, false positives and false negatives, respectively.

### 3. Results

#### 3.1 Data summary

From the final data set of 297 records, there were 184 with passing final grades, and 113 with failing grades. The data set contained records from 77 female students and 220 male students. Student ages were normally distributed, with a peak at the 30-39 age group. The majority (28%) of the students were based in South Australia, followed closely by students in New South Wales (21%). Table 2 shows a summary of final grades per course and its corresponding study period.
3.2 Statistically significant features

As shown in Table 3, a VIF test resulted to 15 demographic and 4 performance-related features after excluding other features sharing high multi-collinearity with the others. A chi-square test applied to the remaining features in Table 3 further revealed statistically significant features (P-value < 0.05) in relation to the target variable Failed. As listed in Table 4, completing the final exam (P-value: 2.15 e-35), submitting all assessments (P-value: 1.48 e-24), and failing to submit any of the marked assessments (P-value: 2.07 e-09) were found to be associated with successful outcome.

Table 3. Summary of features selected via VIF

<table>
<thead>
<tr>
<th>Feature</th>
<th>Type of feature</th>
<th>VIF Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part_Time</td>
<td>Demographic</td>
<td>3.86</td>
</tr>
<tr>
<td>Gender (Male)</td>
<td>Demographic</td>
<td>4.03</td>
</tr>
<tr>
<td>Location (New South Wales)</td>
<td>Demographic</td>
<td>7.02</td>
</tr>
<tr>
<td>Location (Northern Territory)</td>
<td>Demographic</td>
<td>1.82</td>
</tr>
<tr>
<td>Location (Queensland)</td>
<td>Demographic</td>
<td>4.76</td>
</tr>
<tr>
<td>Location (South Australia)</td>
<td>Demographic</td>
<td>8.14</td>
</tr>
<tr>
<td>Location (Tasmania)</td>
<td>Demographic</td>
<td>2.03</td>
</tr>
<tr>
<td>Location (Victoria)</td>
<td>Demographic</td>
<td>5.12</td>
</tr>
<tr>
<td>Location (Western Australia)</td>
<td>Demographic</td>
<td>4.60</td>
</tr>
<tr>
<td>Location (overseas)</td>
<td>Demographic</td>
<td>1.26</td>
</tr>
<tr>
<td>Age (22-24)</td>
<td>Demographic</td>
<td>3.6</td>
</tr>
<tr>
<td>Age (25-29)</td>
<td>Demographic</td>
<td>5.68</td>
</tr>
<tr>
<td>Age (30-39)</td>
<td>Demographic</td>
<td>9.06</td>
</tr>
<tr>
<td>Age (40-49)</td>
<td>Demographic</td>
<td>4.42</td>
</tr>
<tr>
<td>Age (50 and above)</td>
<td>Demographic</td>
<td>2.27</td>
</tr>
<tr>
<td>Completed_Exam</td>
<td>Performance</td>
<td>8.66</td>
</tr>
<tr>
<td>All_Assessment_Submitted</td>
<td>Performance</td>
<td>4.36</td>
</tr>
<tr>
<td>No_Assessment_Submitted</td>
<td>Performance</td>
<td>1.47</td>
</tr>
<tr>
<td>Activity_Count_0</td>
<td>Performance</td>
<td>1.34</td>
</tr>
</tbody>
</table>

Table 4. Summary of statistically significant features selected via chi-square test

<table>
<thead>
<tr>
<th>Feature</th>
<th>Type of feature</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completed_Exam</td>
<td>Performance</td>
<td>2.15 e-35</td>
</tr>
<tr>
<td>All_Assessment_Submitted</td>
<td>Performance</td>
<td>1.48 e-24</td>
</tr>
<tr>
<td>No_Assessment_Submitted</td>
<td>Performance</td>
<td>2.07 e-09</td>
</tr>
</tbody>
</table>

3.3 Model performance

As shown in Table 5, the model trained using only demographic features was able to identify 64 records with a final passing mark correctly and 71 records with a failing final mark correctly. The model trained using only performance-related features was able to identify 182 records with a final passing mark correctly and 76 records with a failing final mark correctly. The model trained using statistically significant features was able to identify 184 records with a final passing mark correctly and all records with a failing final mark correctly. The model trained using all features was able to identify 181 records with a final passing mark correctly and 76 records with a failing final mark correctly.

Table 6 shows a summary of model performance for each of the 4 models. The model trained using only demographic features yielded an accuracy of 45.45%. The model trained using only performance-related features yielded an accuracy of 86.86%. The model trained using statistically significant features yielded an accuracy of 86.53%. The model trained using all features yielded an accuracy of 86.53%
Table 5. Average result form 10-fold cross-validation for each model

<table>
<thead>
<tr>
<th>Model</th>
<th>True Negative</th>
<th>False Negative</th>
<th>True Positive</th>
<th>False Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic</td>
<td>64</td>
<td>42</td>
<td>71</td>
<td>120</td>
</tr>
<tr>
<td>Performance</td>
<td>182</td>
<td>37</td>
<td>76</td>
<td>2</td>
</tr>
<tr>
<td>Statistically significant</td>
<td>184</td>
<td>40</td>
<td>73</td>
<td>0</td>
</tr>
<tr>
<td>All features</td>
<td>181</td>
<td>37</td>
<td>76</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 6. Summary of model performance

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic</td>
<td>37.17%</td>
<td>62.83%</td>
<td>34.78%</td>
<td>45.45%</td>
</tr>
<tr>
<td>Performance</td>
<td>97.43%</td>
<td>67.25%</td>
<td>98.91%</td>
<td>86.86%</td>
</tr>
<tr>
<td>Statistically significant</td>
<td>100%</td>
<td>64.60%</td>
<td>100%</td>
<td>86.53%</td>
</tr>
<tr>
<td>All features</td>
<td>96.20%</td>
<td>67.25%</td>
<td>98.36%</td>
<td>86.53%</td>
</tr>
</tbody>
</table>

4. Discussion

This study demonstrates the ability of limited demographic and performance-related data to predict student outcome in a fully online programming course. Based on the results, demographic features, specifically the ones included here were found to be weak predictors of successful outcomes in learning programming in a fully online environment. This result is consistent with other studies that looked at predicting student learning outcomes using learning analytics for other courses in online learning environments (Hu, Cheong, Ding, & Woo, 2017). It should be noted that the locations included in this study were all in Australia. But with an increase in the number of student enrollments from locations outside Australia, it would be interesting to look at how this can impact the model.

The model performance results, however, reveal how engagement features could predict student success in this type of learning environment. Adding performance-related features to demographic features was able to significantly improve the predictive performance of the trained model from 45.45% to 86.53%, indicating the importance of these data in training such models. Specifically, completing the online final exam and summative assessments were found to be statistically associated with successful outcome in learning programming in a fully online environment. Moreover, training a predictive model based solely on these 3 features demonstrated its reliability in predicting success outcome in this case as supported by an 86.53% prediction accuracy.

It is important to note that student engagement in terms of completing non-mandatory assessments was found to have no statistically significant association with student outcome for fully online programming courses. This result is different from studies that looked at MOOCs general courses and not specific to programming courses (Baker et al., 2015) where early access to resources, constant access to the courses and performing well in the formative non-marked activities are the indicators for the students' successes. There are several factors that explain why the results for the predictors online programming courses are different. First, programming needs constant practice in coding and the number of times a student accesses the courses without "doing the coding" does not affect the student's final grade. Second, in the case of the online students for the programming courses in this study, most students do not usually engage in formative non-marked learning activities and usually only view this as additional workload. This is the reason why continuous summative assessments were introduced in these courses for UniSA Online. Since non-mandatory assessments were not accounted for in the computation of the final grade, it is understandable that this shows no statistically significant association with passing the course. Lastly, with regard to other engagement methods, the online programming courses used in the study only use discussion forums primarily as a tool for asking question. This is similar to how MOOCs EdX programming courses use forums (Waller, 2019). Because of this, it is unknown if engagement in the forums is a good predictor of a successful outcome in the course. For future studies, text analysis of the discussion forums can be investigated if it can help identify students at-risk.

For online learning facilitators, this study reiterates the need to monitor student engagements in the submission of summative assessments (Baker et al., 2015). This also informs how learning facilitators can adjust how they monitor groups of students, especially those at risk of failing. For
instance, additional reminders can be set to ensure students do not miss submitting assessments and completing exams required for the course. By improving monitoring check-ins especially before assessment due dates, early intervention for non-submission of initial assessments can be prevented.

It should be noted that this study is not without any limitations. First, the features included were limited to general demographic data (e.g., name, age, gender location, type of study). Analysis of other factors such as employment status, type of work, and basis of entry can be added to further see if demographics is a possible predictor of successful outcomes in an online environment, especially for programming courses. Gender data is still presented as only two binary choices, and it is possible that a third option may have been present but not shown in the reported data. Second, the performance features included in this analysis requires completion of the whole course before a reliable prediction could be made using the models. Therefore, the models cannot be utilized prior to the course being run or during the early phase of the course. Other data that can be used in future studies is looking at the student’s personality (self-efficacy, self-regulation) and previous performances from previous online courses. It is also important to investigate the generalizability of these models for other courses and online learning environments. This study advocates for the collection of finer and more specific student demographic, personality, academic, and behavioural data in a fully online learning environment to enable prediction of success outcomes early on.

Acknowledgements

The authors would like to acknowledge the UniSA Online programming students and online course facilitators for the data used in this study.

References


presented at the Proceedings of the Seventh International Learning Analytics & Knowledge Conference.


LiveDots: Real time Interactive Braille Music Translator to Integrate Blind Students into Music Classes

Claudia GUERRERO-GARCÍA-HERAS*, Óscar DÍAZ-RIBAGORDA*, Lucas DE-TORRE-BARRIO*, Borja MANERO* & María GUIJARRO-MATA-GARCÍA*

*Complutense University of Madrid, Spain
*claudiaggh@gmail.com, bmanero@ucm.es, mguijarro@ucm.es

Abstract: Music is a universal cultural expression; however blind students may have difficulties to follow a sighted music class due to the lack of a common written music language between teacher and student. This problem is also present in other academic fields, like science. An approach to solve it is EDICO, a real time interactive Scientific Editor. We want to take the solution of real time interactivity to the music field. This paper presents the development process of a real time interactive solution called LiveDots and an experiment to test it in which we develop a real time interactive solution called LiveDots and test it with blindfolded users to check if it could help integration of a blind student in a music classroom. Results show that LiveDots was usable by both sighted and blindfolded users and that real time interactivity may be useful to integrate blind students in music class. These results open up a new horizon of solutions based on real time interactivity to ease blind and sighted students’ integration in the same class.

Keywords: Accessibility, blind people, tiflotechnology, inclusive education, education, music, real time interactivity, Braille, Braille music, screen reader, Braille line, computer application.

1. Related work

Braille is the most extended method of tactile reading and writing amongst the blind. It is based on different combinations of embossed dots (Kent, 2012). The extension of Braille to the music field is known as Braille music (de Candé, 2002). Musical Braille not only gives blind students a tool to understand and express music but also shapes how we think and talk about music and, by extension, how we analyse it (Abramo & Pierce, 2013; Johnson, 2015).

However, blind students do not often receive the education needed to understand musical Braille notation and most schoolteachers do not know Braille neither how to facilitate the student's learning. The alternative for blind students in order to improve their musical skills is to attend a school for the blind. In this scenario, blind and non-blind students are taught using different teaching strategies leading to a dichotomy between music Braille and conventional music writing. The lack of familiarity with conventional music writing makes it more difficult for blind students to later join in an environment with sighted musicians, for example in music college or in an orchestra (Goldstein, 2000).

There is a need for blind students to be able to follow a music class with mostly sighted students while learning Braille music notation and understanding the conception of print music (Quaglia, 2015; Buhagiar & Tanti, 2011).

Some of the conventional strategies and tools used by blind or visually impaired students to facilitate participation in class are: enlarged print notation, fellow class members, parents or teachers reading the scores for them and use of Braille music notation whenever it is possible (Frederick & Moss, 2009; Smaligo, 1998). All these tools need either the help of a person who transcribes the score orally or a teacher who knows musical Braille and teaches it to the student which may not always be possible. This makes the student dependent on people around him.

The advance of technology in recent years has given the chance to ease the integration of blind students. There are projects like Braitico (ONCE, n.d.), an inclusive Braille literacy method developed by the ONCE (National Organization of the Blind in Spain) intended for children to learn Braille. In the
field of music, there are computer programs designed to enable visually impaired people to view and edit music in Braille notation (Homenda, 2008) like Braille Music Editor (Veia Progetti, n.d.), studies about how to teach Braille music to children using computer applications (Nicotra and Quatraro, 2008; Borges y Tomé, 2014), approaches to teach music to blind studies using talking music instead of Braille (Capozzi, Prisco, Nasti & Zaccagnino, 2012), score translation programs from print music into Braille music and vice versa like FreeDots (Repain et al, n.d.) and there is even a standard to share Braille music notation in the web called Braille Music Markup Language (Encelle, Jessel, Mothe, Ralalason & Asensio 2009).

These applications allow the teacher to create a print score with an editing score program, like MuseScore (Schweer, n.d.), and to translate it into Braille using a translation program so the student can read it. However, interactivity is missing: if the teacher wants to modify an element of the score to correct it or to make an improvised exercise, first, the changes should be done in the editing score program, then exported and translated into Braille with the translation program and finally the student could see the changes. This is a slow process that does not integrate a blind student into the usual development of a music class.

The problem of integrating blind students in the classroom is present in almost every field in education. There are many different approaches for a solution, like Aim-Math, an interactive-enhanced mathematics learning system for blind and visually impaired students using text-to-speech to read aloud math expressions (Naruedomkul, 2013). However, this approach is not friendly for a class of blind and non-blind students. Another approach is EDICO (Scientific Editor ONCE) (Carenas, Cabra, Mata-García, Gea & Hernández, 2018), a project promoted by the ONCE and developed in cooperation with the University Complutense of Madrid. It is an accessible mathematics, physics and chemistry editor. It translates scientific language in real time from printed writing into Braille and vice versa. This allows blind students to follow a science lesson interacting in real time with a teacher who doesn’t know Braille.

2. Introduction to the project

After the success of EDICO, the ONCE wanted to develop similar solutions to integrate blind students in other subjects, like music. This is how the idea of LiveDots was born: a music editor that translates music scores from print music to Braille music in real time. LiveDots is an innovative desktop application which allows users to read a score in Braille and in print at the same time and to modify the print score and see the changes in the Braille score in real time. This would give the blind students a solution to the lack of interactivity we explained before.

The main objective of this study was to check if the application LiveDots is useful to include a blind student in a music class. We believed that being able to translate scores into Braille in real time would increase the interaction between teacher and blind students improving blind students' participation. This can be verified by checking if people find the application useful and if they would recommend it. To prove our theory, we formulate the following research question:

**RQ1. Can the application LiveDots help with the inclusion of a blind student in a music class?**

In addition, to test the utility of the application LiveDots, it is important to verify its usability. Therefore, another goal of this test was to verify usability for both target users: music teachers and blind music students.

On the one hand, it is required that a blind person can use the application LiveDots without any assistance for the application to be useful. It is only necessary to check that the blind person can perform the basic functionalities such as opening and reading a score. When a person is able to read the score, no matter if he or she is sighted or blind, it directly follows that he or she can read the changes when made. Thus, the keyboard shortcuts and the screen reader must be practical and easy to use. To check this, we posed the following research questions:

**RQ2. Can a blind person use the application LiveDots without any assistance?**

On the other hand, a teacher should be able to use LiveDots even without knowledge on musical Braille. The actions loading a score and doing score changes have to be uncomplicated. Therefore, the last research question was:

**RQ3. Can a sighted person with musical background who doesn’t know musical Braille use the application LiveDots?**
3. **LiveDots application characteristics**

Building on the previous requirements (a teacher should be able to use the application without any knowledge about musical Braille and blind students should be able to use the application without any assistance), we decided to implement the following features in LiveDots (Figure 1).

The application uses scores in MusicXML format since it is the most used musical representation format. When it is selected, the score is showed both in print and in Braille. The print score for sighted users is displayed by a stave and musical elements and the Braille score is shown in Figure 1.

Musical Braille notation to be read with a refreshable Braille display. On one side, a sighted user can edit the print score. These changes are shown in real time in the musical Braille score, so a blind student could read the new score in the refreshable Braille display at the same time it is modified. On the other side, a blind user can use keyboard shortcuts to use the application LiveDots and read the Braille score by a Braille line or using the screen reader: when placing the focus in the musical Braille score displayed on the screen, the screen reader will say the musical elements as you go through them. These features: screen reading of Braille musical elements and real time modification are an innovation of the application LiveDots. Other applications allow you to edit the score (for example, Musescore), but not in real time, or can be used with a screen reader (for example, Braille Music Editor), but they do not say each musical element in the Braille music score when you select them.

To deal with the translation we took inspiration from a study made on the matter (Goto, D., Minamikawa-Tachino, T. and Gotoh, N., 2007) but in order to allow the real-time translation that makes our application unique we used data binding combined with event handling. The other innovative feature, the Braille score reading, was accomplished by programming a personalized script for the screen reader (JAWS) to use with our application.

Lastly, the application LiveDots can be used by reduced visibility people and color-blinded people. The application allows to regulate the zoom (both in the scores and in the menu) and it is compatible with the Windows Colorblind Mode.

4. **Methodology**

4.1 **Participants and experimental design**
The study involved 7 participants. All of them were sighted people without any knowledge on Braille or Braille music. Each of them tested the application twice. They did it first blindfolded and then watching the application. Lastly, they did a questionnaire about LiveDots.

The experiment was designed in order to test if LiveDots could help to include blind students in a music class. First, the participants checked the application and then they gave us feedback with a questionnaire (Figure 2). For the experiment, each participant was provided the application LiveDots. In addition, they used the screen reader JAWS.

4.2 Questionnaire

The questionnaire had two initial yes/no questions. The first one was if getting Braille scores with informatic applications is easier than the traditional way of printing it on a piece of paper, and the second one was if the participant knew an application of translation or reading Braille scores different from LiveDots. Our hypothesis was that the answers would be yes and no, respectively. This would highlight the need of an application like LiveDots to do this translation and reading supporting the first research question (RQ1).

After these two questions, the questionnaire consisted of three different parts to check: usefulness, usability by blind people and usability by sighted people. They consisted of 5 questions each valued in Likert 5 scale, being 1 the worst scoring for the application LiveDots and 5 the best punctuation. To calculate the average value of each question we calculated the mean of the values of the responses.

Finally, there was an open question about what improvements the participant would make to the application LiveDots.

5. Results

After conducting the experiment, we obtained 7 answers for the blindfolded test questionnaire. We obtained a low answer rate due to the unexpected worldwide pandemic situation which did not allow us to conduct the experiment in the ONCE facilities. Having said that, we managed to get and analyse the results the best way possible. For the Likert 5 questions we calculated the mean of the answers obtained for each question and the results are shown in the figure 6.

For each group of related questions, we calculated the arithmetic mean of the average answers and represented it in a box plot (Figure 3). For the yes/no questions we obtained that 100% of the testers think that the use of computer applications makes obtaining a score in Braille easier than the traditional methods. And 85.7% of the testers don’t know any computer application for translation and/or reading of braille scores.
6. Discussion

Throughout this section, we will try to answer the research questions set out in section 2 by discussing the evidence reported in the results section. First, we note the need of an application like LiveDots in the fact that although all of the participants (100%) think that the use of computer applications makes the task of obtaining Braille scores easier than the traditional method of printing on paper, hardly any of them (14%) know an application to accomplish it. As to test the utility of the application LiveDots (RQ1) it is necessary to test usability first (RQ2, RQ3), we will answer the research questions in that order.

**RQ2. Can a blind person use the application LiveDots without any assistance?** Yes. The blindfolded test results show that a blindfolded person can successfully perform the main tasks available in the application (4.36/5 in blind usability B1-B5).

**RQ3. Can a sighted person with musical background who doesn’t know musical Braille use the application LiveDots?** Yes. The sighted test results show that participants without any knowledge on Braille music could use the application without any trouble (4.33/5 in sighted usability S1-S4). Indeed, most of the participants (85%) think that little to no knowledge on Braille music is needed to use the application.

**RQ1. Can the application LiveDots help with the inclusion of a blind student in a music class?** Yes. After performing the blindfolded and sighted test the participants think that the application and its functionalities can be useful for integrating a blind student in a music class (4.38/5 in usefulness U1-U3) and they would probably use and recommend it (4.08/5 in P1-P2). In addition, RQ2 and RQ3 show that the developed application is also usable, which is a needed requirement for it to be useful.

7. Conclusions and future work

We developed LiveDots as an application to ease the communication between blind music students and sighted music teachers. The present study about provided empirical evidence that real time interactive applications can be used to integrate blind students in sighted music classes. In fact, LiveDots is an example in the music field that can be used to improve accessible education in our society.

The first conclusion is real time interactive applications, such as EDICO (application of Mathematics), can be generalized to artistic fields. Also, LiveDots design and the proposed experiment confirmed that the way blind people use computers is entirely different from how sighted users do it. Blind people use a screen reader and the keyboard to navigate with ease and use shortcuts repeatedly. On the other hand, sighted people typically use the mouse to navigate and do not usually know the place of keys.

This study had certain limitations. The instruments used were limited. The blindfolded experiment with sighted people gives an insight into answering our research questions, however the limited time available for the experiment and the global pandemic situation did not allow us to test the application with blind people. Non-sighted people could test the application deeper and check if habitual accessible procedures are followed in the application and if it is comfortable to use.

As a result of the success of the study, we are preparing a new experiment in cooperation with the ONCE to take this technology to the classroom. This will be a study involving blind people and music teachers to check the potential of the application LiveDots.
Acknowledgements

This project has been partially founded by the Ministry of Science, Innovation and Universities of Spain (Didascalias, RTI2018-096401-A-I00) and it was supported by the ONCE-Tiflotechnology Chair. We would also like to thank to the ONCE for their collaboration in this project, particularly to Pablo Carreño Gea, who was part of the EDICO project, for his cooperation in the development of LiveDots.

References


Learner Model of Knowledge Grounding in Discovery Learning

Takuya WADA*, Yuki HAYASHI*, Kazuhisa SETA* & Tomoko KOJIRIb

a Graduate School of Humanities and Sustainable System Sciences, Osaka Prefecture University, Japan
b Faculty of Engineering Science, Kansai University, Japan
* ddb04005@edu.osakafu-u.ac.jp

Abstract: Discovery learning is a learning method for acquiring new knowledge and utilizing such knowledge in the real world. As a framework for support discovery learning, several exploratory learning environments (ELEs) have been developed, which aim to acquire and utilize knowledge within such environments. However, some learners cannot utilize their knowledge because it is weakly linked to entities in the real world. Few ELEs aim at using learning to strengthen the connection of the learner’s knowledge with real-world entities. In view of the insufficient framework for supporting a discovery learning that links knowledge with entities, this study identifies the cognitive state of errors in the grounding of discovery learning, and internal representations of information necessary to detect such errors in an ELE.

Keywords: discovery learning, learner model, exploratory learning environment, grounding, biology learning

1. Introduction

In recent years, various types of active learning have been introduced in primary and secondary school education, among which discovery learning is one of them. The purpose of discovery learning is to acquire new knowledge and to utilize such knowledge in the real world. Choi et al. (1995) stated that some learners are unable to utilize such knowledge in everyday circumstances, and that activities utilizing knowledge in the real world should be emphasized. Knowledge in the field of natural science mainly refers to matter or phenomena in the natural world. If a learner’s knowledge is not connected (hereinafter called “grounded”) with matter or phenomena in the natural world as entities, the learner may not be able to utilize it. For learners to be able to utilize knowledge in the real world through discovery learning, a type of learning that strengthens the learner’s grounding with real-world entities is necessary. This study considers knowledge as concepts and their relationships, and entities as external resources that compose the real world. Grounding is regarded as a state in which the knowledge of the learner is connected with the correct entities.

De Jong et. al. (1998) stated that learners should have the knowledge beforehand if discovery learning is to be fruitful. Utilization of this knowledge is a prerequisite for improving the quality of the learning outcomes. However, applying knowledge to the learning targets requires their appropriate grounding. If the grounding is incorrect, the knowledge cannot be used effectively. For example, to identify the species of observed unknown plants, the learner needs to utilize knowledge of characteristic plant parts to identify them. However, if the learner cannot ground plant parts of the knowledge with parts of the unknown plant, the knowledge is not utilized well. As a framework to support discovery learning, several exploratory learning environments (ELEs) have been developed. Santos et al. (2012) stated that the interaction of learners is highly unstructured and is difficult to model and prescribe. They proposed an approach to develop intelligent support in ELEs. However, a specific way to identify the status of a learner’s grounding and types of log data as useful information for the identification has not been proposed.

Because a general method for evaluating whether a learner can apply knowledge to an entity is yet to be clarified, the cognitive state that indicates the accuracy of the grounding is unknown. Therefore, this study deals with the question of what mechanism can capture a cognitive state through which knowledge can be correctly applied in the real world. This study identifies a cognitive state of
errors in the grounding of discovery learning, as well as internal representations to detect such errors from interactions between an ELE and a learner.

2. Related Studies

Conventional ELEs promote the utilization of knowledge by providing a virtual learning environment that does not cause grounding errors. Table 1 shows characteristics of each type of ELE. Simulators and microworlds are virtual worlds generated on a computer with the aim of encouraging learners to discover specific knowledge through direct operations and observations. Simulators offer a virtual world as a representation of a domain model. A flight simulator is an example of a simulator, as indicated by Perry (2004). Microworlds represent the domain model and allow learners to add or remove elements. As an example of a microworld for physics learning, a virtual environment in which learners can change the conditions of a moving object and its surroundings and observe their behavior, imitating the real-world, was developed by Cockburn et al. (1995). In general, simulators and microworlds cannot assess a learner and do not have a diagnostic function for intelligent support.

An intelligent microworld is a learning environment using a learner model and intelligent support functions to help learners who are struggling with using a microworld. There are several learner modeling techniques applied in an intelligent microworld. An intelligent microworld for chemical experiments developed by Yoshikawa et al. (2000) records the operational sequence of the learners and estimates their goals as well as the errors in their operating procedures and knowledge. In addition, a physics-based intelligent microworld developed by Reid et al. (2003) estimates laws that the learner is unaware of as based on experimental results and the derived laws inputted by the learner.

An error-based simulation (EBS), which is a method for generating phenomena by applying erroneous concepts used by students, is a promising method allowing the students to be aware of errors based on certain correctly known phenomena connected to their proper concepts. EBS promotes the correction of learner errors in knowledge by providing an unreal world that expresses the learner's erroneous knowledge as entities. Because the learning target is a correction of specific knowledge errors, EBS is introduced in a learning context with explicit tasks. Hirashima et al. (2009) defined the drawing of a force direction based on errors in the learner’s knowledge, and their developed EBS system estimates errors in the learner’s knowledge from drawings inputted by the learner.

Table 1

<table>
<thead>
<tr>
<th>Characteristics of several types of ELE</th>
<th>ELE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>simulator</td>
</tr>
<tr>
<td>learning task</td>
<td>given</td>
</tr>
<tr>
<td>intelligent support</td>
<td>without</td>
</tr>
<tr>
<td>learning target</td>
<td>knowledge acquisition</td>
</tr>
</tbody>
</table>

Most ELEs target the acquisition of knowledge, and there are few ELEs that target the correct grounding of the learner. Although some ELEs support error correction of knowledge to reach the correct grounding such as EBS developed by Hiramoto et al. (2018), few models exist that capture cognitive states by distinguishing knowledge errors from grounding errors. In such ELEs, although the learner can temporarily apply their knowledge to only entities provided by ELEs, they may not be able to apply their knowledge correctly in the real world, which allows grounding errors to occur. For example, suppose that the learner applies knowledge of angiosperm to an entity of a plant that the ELE shows as an angiosperm. Then, even if the learner observes other angiosperms in the real world, the learner may not be able to apply their knowledge of an angiosperm to these entities. In this case, the correct application of knowledge requires having knowledge regarding the characteristics of the
angiosperms and a correct grounding of such knowledge and the characteristics of the entities. The learning target focused on in this study is a grounding of knowledge and entities in virtual worlds on computers, and a learner modeling method for detecting grounding errors is established. This study focuses on complex cognitive states in a learning context in which the learner can decide on a learning task and observe numerous targets, identifying the cognitive states of such errors.

3. Types of Errors of Grounding and Representation Method

3.1 Errors of Grounding

This study regards knowledge as a binary relation between two concepts and focuses on hierarchical knowledge structures as complex knowledge structures. The target knowledge structure in this study consists of “is-a” relationships, which indicate inheritance relationships between concepts, and “has” relationships, which indicate ownership relationships between concepts, and covers the domains of the plant types and their parts. Because discovery learning for many students is carried out after learning each concept of the domain as prior knowledge, learners who have formed at least individual concepts of a domain knowledge structure correctly are targeted. Although these learners have individual concepts, the relationships between their concepts may be incorrect.

Correct grounding is a state in which each knowledge concept is applied to the correct entities, and the correct relationship of the knowledge is applied to the relationship between entities. Errors of grounding during learning with errors of knowledge and grounding are as follows:

i. The relationship between concepts of a knowledge structure is incorrect.
ii. The combination of the applied concept of a knowledge structure and an entity is incorrect.
iii. The relationship between entities captured by the learner is incorrect.

Error I is a state in which the learner cannot properly apply knowledge to the correct entity because the knowledge is incorrect. For example, a learner who mistakenly believes that a “pine tree has coconuts” cannot be grounded correctly because the learner has a discrepancy in the observation result that “pine trees do not have coconuts” when observing the pine tree as an entity. Error II is a state in which individual concepts are applied to incorrect entities. For example, the learner recognizes a palm entity as a pine tree or recognizes a coconut entity as a conifer cone. Error III is a situation in which the learner overlooks the entity and misunderstands the relationship between different entities. For example, the learner recognizes that “the palms do not have coconuts” when observing a palm entity and overlooks the coconut. In this case, even though the learner has the correct knowledge that “a palm has coconuts,” the learner cannot ground the knowledge correctly owing to a discrepancy between the ownership relationships of the knowledge and entities. These individual errors occur concurrently. To define the state of the learner in which these errors occur, this study proposes a representation method of the cognitive state of the grounding of knowledge and entities. Furthermore, the state of the learner defined based on the representation method is formulated in a computer readable manner.

3.2 Representation of Errors of Grounding

As a learner’s concept system in discovery learning, we consider concepts acquired during classroom learning and concepts formed through observation. This study calls the former a “knowledge layer” and the latter an “observation layer.” Species of plants include individual plants (e.g., pine trees and palms) and groups include multiple plants (e.g., seed plants and gymnosperms). There are inclusive relationships among them. Species of plants have parts in common (e.g., stamens and calyces), and there are ownership relationships between species of plants and parts of plants.

Knowledge and observation layers can be represented as a concept map of plants and their constituent parts. In the concept map, two types of nodes representing “species” and “parts” of plants are introduced, and the inclusion relationship of the plant types is represented through an “is-a” relationship link, whereas the ownership relationship between concepts is represented as a “has” relationship link. A specific concept of species inherits the “has” relationships of its generic concepts. In addition, “not has” relationships are introduced in the observation layer to explicitly express a cognitive state in which the learner is aware of a nonexistence of a part of an observed plant. This study
approximates a state in which relationships between a concept of knowledge layer and an entity of observation layer and relationships between relationships in both layers are correct as correct grounding. These relationships are represented as “instance-of” relationship links. The correspondence between links of both layers is also represented as an “instance-of” relationship link because relationships between concepts are also observed as relationships between entities.

Figure 1 shows an example of the proposed concept map. The entity connected to the concept “XXX” by the “instance-of link” is described as “#XXX”. In the knowledge layer, is-a links indicate that “pine trees and palms are gymnosperms,” and “has links” indicate that “gymnosperms, pine trees, and palms have male flowers,” “pine trees have pineapples,” and “palms have coconuts.” The pine trees and palms in the knowledge layer inherit the “has link” of male flowers from the gymnosperms. In the observation layer, “a pine tree has male flowers” and “a pine tree does not have coconuts” are recognized through observation. Regarding an “instance-of” relationship between the relationships of two layers, which is omitted in Figure 1, the correspondence of the “has link” between “pine trees” and “male flowers” in the knowledge layer and the “has link” between the “#pine tree” and the “#male flower” in the observation layer indicates the correct instance-of relationship. By contrast, because the “has link” between the “pine trees” and “pineapples” in the knowledge layer does not correspond with the “not has link” between the “#pine tree” and the “#pineapple” in the observation layer, this instance-of relationship between the relationships of both layers is incorrect.

Error I corresponds to an error of a link in the knowledge layer. Error II corresponds to an “instance-of link” between both layers. Error III corresponds to an error of a link in the observation layer. The types of errors in these relationships are as follows:

a). The “has link” in the knowledge layer is incorrect (Error I).
b). The “is-a link” in the knowledge layer is incorrect (Error I).
c). The “instance-of link” is incorrect (Error II).
d). The “has / not has link” in the observation layer is incorrect (Error III).

In the binary relationship between the knowledge layer and the observation layer, the knowledge and entities are grounded correctly when “instance-of links” between concepts and the “instance-of link” between “has links” are correct. A state in which these “has links” and the “instance-of link” are correct indicates that the “instance-of” relationship between the relationships of two layers is also correct. Therefore, an error in grounding is exposed as an error of the “has link” in the knowledge layer, an error of the “instance-of link” between concepts, or an error of the “has / not has link” in the observation layer. Although an error of the “is-a link” is not exposed as a grounding error, the “has link” incorrectly inherited by the “is-a link” is exposed as such an error.

![Figure 1. Example of the proposed concept map](image)

4. Learner Modeling of Grounding State

This study focuses on a discovery learning in which the domain has a knowledge structure consisting of inclusion relationships and ownership relationships between concepts, and proposes a learner model that expresses grounding errors, as described in Section 3. The domain has a correct knowledge
structure with multiple hierarchies of inclusion relationships and single hierarchies of ownership relationships. The correct concept system consists of the knowledge structure of the domain and the corresponding entities are defined. To consider “instance-of links” between the learner’s concepts and “instance-of links” between “has links,” the learner’s partial grounding state is expressed by understanding the state of the binary relation of the knowledge layer and the corresponding binary relation of the observation layer of the correct concept system.

Figure 2 shows a part of the correct concept system corresponding to the learner’s partial grounding state. “species X” is any concept in a correct concept system, and the part that “species X” has is described as “part x.” In addition, “#species X” is described as the correct entity of “species X,” and “#part x” is described as the correct entity of “part x.” By integrating the partial grounding states, the learner’s grounding states for the overall correct knowledge system is modeled.

![Correct concept system diagram](image)

**Figure 2.** Part of the correct concept system corresponding to the learner’s grounding state

Discuss internal representations to detect errors in Section 3. Table 2 shows a learner model that summarizes attributes and values that the computer must capture to detect each error.

<table>
<thead>
<tr>
<th>attribute</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>a1. Existence of “has link” between “species X” and “part x” in the knowledge layer</td>
<td>true, false</td>
</tr>
<tr>
<td>a2. Existence of “#species X” in the observation layer</td>
<td>true, false</td>
</tr>
<tr>
<td>a3. Existence of “#part x” in the observation layer</td>
<td>true, false</td>
</tr>
<tr>
<td>a4. State of “has link” of “#part x”</td>
<td>has, not has, null, with other</td>
</tr>
<tr>
<td>a5. State of “instance-of link” of “#species X”</td>
<td>true, false, null</td>
</tr>
<tr>
<td>a6. State of “instance-of link” of “part x”</td>
<td>true, false, null</td>
</tr>
<tr>
<td>a7. State of “instance-of link” of “#species X”</td>
<td>true, false, null</td>
</tr>
<tr>
<td>a8. State of “instance-of link” of “#part x”</td>
<td>true, false, null</td>
</tr>
</tbody>
</table>

To detect error a, it is necessary to obtain the information about the “has links” in the learner’s knowledge layer. Error a is a state in which the “has link” of the knowledge layer of the correct concept system is not in the learner’s knowledge layer. In addition, the state in which the “has links” that are not in the correct concept system exists in the learner’s knowledge layer is also error a. Therefore, the learner model requires the attribute of existence of the “has link” between “species X” and “part x” in the knowledge layer. For the value of this attribute, “exist” is described as “true” and “not exist” is described as “false.”

Error c is a state in which there is no “instance-of link” between correct concepts of both layers, and these concepts have an “instance-of link” with the incorrect concept. To detect error c, the learner model requires the attributes indicating the state of the “instance-of links” for each species X, “part x,” “#species X,” and “#part x.” For the value of these attributes, “correct” is described as “true,” “incorrect” as “false,” and “not existence” as “null.” The error of “instance-of link” between the “has links” of two layers is represented by the error of the “has link” in knowledge or observation layer.

To identify error d, information of the existence of “#species X” and “#part x” in the observation layer and “(not) has link” between these concepts is required. Error d is a state in which “#species X” and “#part x” are in the observation layer, and these concepts have a “not has link.” The
state in which “#part x” in the observation layer has a “has link” with a concept other than “#species X” is also error d. Therefore, to detect such errors, the learner model requires the attributes of existence of “#species X” and “#part x” in the observation layer. For the value of this attribute, “exist” is described as “true,” and “not exist” as “false.” Furthermore, an attribute that indicates the state of “has link” of “#part x” is also required. The value of this attribute is described as “has,” “not has,” “null (i.e., no link), and “with_other (i.e., “#part x” has a “has link” with a concept other than “#species X”).”

The learner's grounding state can be approximated by obtaining values of all attributes in Table 2 for all binary relation of the overall knowledge and observation layer. This modeling assumes that the ELE is designed such that values of all attributes in Table 2 can be obtained from interactions between an ELE and a learner. By updating the learner model every time the learner performs a specific operation, the grounding state of the learner can always be captured.

5. Conclusion

This study focused on discovery learning for strengthening the grounding between knowledge and entities in the real world, for which a framework for learner modeling was proposed. Specifically, the types of errors in the grounding of discovery learning were identified, and the information necessary for a computer to detect such errors was suggested. These errors are formalized into computer-readable attributes and values, and the learner model for ELE that can obtain such information was proposed.

Currently, a knowledge structure with single hierarchies of ownership relationships is being targeted. However, as an example, some plant parts also have its more detailed parts in the real world, and therefore, in a further study a learner model for a knowledge structure with multiple hierarchies of ownership relationships should be developed.

In some cases, learners may be able to correct a grounding error themselves. Although the proposed learner model can capture individual errors in the grounding of the learner, this model cannot determine whether a learner is struggling from such grounding errors. The state of errors in the grounding and knowledge indicating the struggles of a learner must be identified, and a learner model that can capture them should be developed.

Acknowledgements

The work was supported in part by JSPS KAKENHI (No. 19H01725).

References

Using the Community of Inquiry framework to develop an educational chatbot: lesson learned from a mobile instant messaging learning environment

Congyao QIN, Weijiao HUANG* & Khe Foon HEW
Faculty of Education, The University of Hong Kong
wjhuang1@connect.hku.hk

Abstract: Chatbots can be defined as artificial narrow intelligence (ANI) that is programmed to perform a single task such as answering customers’ questions by harnessing the power of machine learning. However, there is a lack of educational-related chatbots developed to interact with learners in mobile instant messaging (MIM) apps, as well as a lack of theoretical grounding for chatbots for teaching and learning purposes. In this study, a chatbot taking on the role of an instructor, was developed to teach adult students fake news identification in an online course via the WeChat MIM app. We provide a detailed description of the MIM chatbot development guided by the Community of Inquiry theoretical framework. We present the evaluation of the MIM chatbot based on a self-reported survey regarding students’ perceptions of the teaching presence, social presence, and cognitive presence. The findings revealed social presence mainly existed between the chatbot and individual student in a MIM learning environment. We hence offer some preliminary design principles for future educational chatbot design.

Keywords: Chatbot, mobile instant messaging, WeChat, artificial narrow intelligence

1. Introduction

Chatbots have made inroads into the educational environment with its capability of interacting with students in natural language. The common ways of implementing chatbots include chatbot taking the role of teacher to delivery instructional materials (Heller, Proctor, Mah, Jewell, & Cheung, 2005), and chatbot playing the role as a learning partner to facilitate learners’ understanding of the subject matter (Fryer, Ainley, Thompson, Gibson, & Sherlock, 2017). Presently, many chatbots have been developed to interact with users on standalone platforms, rather than through mobile instant messaging (MIM) apps (Schmulian & Coetzee, 2019). Educators can make use of these applications to build up an accessible communication channels with students (Tang & Hew, 2017). Unlike mobile apps which need to be downloaded and take up valuable space on a user’s phone, chatbots in MIM are readily available within a MIM app and they do not need to be downloaded (Schlicht, 2016). Yet, despite the significant potential for automated learning, there is very little research exploring the use of educational chatbots in MIM apps (Schmulian & Coetzee, 2019). The purpose of this study, therefore, is to fill this research gap. We aim to provide a detail description of the development of a MIM educational chatbot, using a visual chatbot development platform Dialogflow that does not require computer programming knowledge. The theoretical rationales employed in this study is based on the Community of Inquiry (CoI) framework. A two-round case study was conducted to investigate students’ experience in learning with the MIM chatbot. The effects of the CoI framework the MIM chatbot design were measured by students' self-reported surveys to answer our research question: What are students’ perceptions of the teaching presence, social presence, and cognitive presence afforded by the MIM chatbot?

In the following sections, we will present (a) the theoretical background of our work, (b) an instructional design of MIM chatbot with the CoI framework, (c) MIM chatbot evaluation based on students’ self-reported surveys, and (d) a set of preliminary educational chatbot design principles.
1.1 Chatbot-assisted Learning

Chatbot is a dialogic system that can interact with users in natural language. Unlike a human employee, chatbots can always offer help on demand 24/7 (Garcia, Fuertes, & Molas, 2018). Early chatbots were created by using the Artificial Intelligence Markup Language (AIML), a type of XML programming language (Kane, 2016). With the help of visual development platforms, programming knowledge is not a must-have skill to design a chatbot. Via built-in Application programming interface (API) keys generated by the development platforms, chatbots can be handy for non-technical users. Chatbots show the pedagogical value in facilitating learners’ language knowledge acquisition (Ruan et al., 2019; Wang, Petrina, & Feng, 2017), and fostering learners’ willingness to communicate (Ayedoun, Hayashi, and Seta, 2019).

Chatbots interacting with users in a MIM app may offer educators an automated means of content instruction (Nakpodia, 2017). Carayannopoulos (2018) examined a messaging chatbot which was designed to help first-year students to adapt to university life. The chatbot in his study is as same as any friend in students’ contact lists that students can seek information from the chatbot in the way of daily communication. The findings indicated that students gained a sense of being connected with their professors. Gengobot, a Japanese grammar dictionary chatbot integrated into the messaging application LINE, enables personalized learning for students to practice basic grammar knowledge (Haristiani & Danuwijaya, 2019). Similarly, Schmulian and Coetzee (2019) developed two Facebook Messenger bots to instruct university students in a large accounting class and assist students in understanding learning materials at their own pace. Chatbots merged with messaging applications can offer students a livelier and friendlier learning experience.

1.2 The Community of Inquiry as a Theoretical Framework

The Community of Inquiry (CoI) framework is initially intended to build a constructivist learning environment in computer conferencing and online learning in higher education context (Garrison, Anderson, & Archer, 2000). After nearly two decades, this theoretical framework has been applied to diverse educational settings, including face-to-face learning and K-12 learning (Garrison, 2016). The CoI framework is comprised of three constructs, namely social presence (e.g., personal, open communication, and group cohesion), cognitive presence (e.g., triggering event, exploration, integration, and resolution), and teaching presence (e.g., design and organization, facilitating discourse, direct instruction), to articulate the knowledge construction in online learning community. Social presence refers to students’ the ability to present themselves as real in a community of inquiry with keeping their individual personalities (Garrison et al., 2000). Cognitive presence is defined as the extent of learners’ information construction through the sustained interaction and reflection in the community. Teaching presence includes designing curriculum, giving direct instructions and facilitating disclosure (Garrison et al., 2000). A personal meaningful communication and learning experience can be offered to students by using the above three presences to develop learning contents. On that account, we designed a MIM chatbot guided by the CoI framework.

2. The Development of Fake News Bot

2.1 Chatbot Design with the CoI framework

We designed a chatbot as a role of online instructor to teach university students how to identify fake news. Table 1 presents the information in learning objectives and students’ learning materials. Students are expected to interact with the chatbot to recognize the features of fake news, figure out fake news by using a four-step method and apply the methods in real life. The training data of chatbots were designed by the instructor with pre-set prompts according to the learning materials.

The learning activities of Fake News bot were grounded on the categories in social, cognitive, and teaching presence (Table 2). Social presence was visualized by the help of emojis to express the feelings of chatbot during the student-chatbot interaction and form a risk-free learning environment. For instance, the chatbot greeted students with a smiley face emoji and showed agreements with a thumb up
emoji. Vocatives were also merged to facilitate a collaborative climate. For example, chatbot called student’s name at the beginning of the responses. Cognitive presence was presented with diverse teaching clues. When an example of fake news was given to students, the chatbot asked students whether the news is fake or real (triggering event). After students typed the answers, chatbot would give immediate feedback and explain the features of fake news through a dialogue way (exploration). Then, several cases were delivered to help students recall the factual knowledge of fake news (integration). Once students finished the knowledge construction via the interaction with the chatbot, additional learning materials would be given to let students apply the information in a new situation (resolution). Teaching presence was demonstrated by having an intro-video to illustrate learning outcomes clearly and using short video as knowledge instruction.

Table 1. Summary of Learning Objectives and Learning Materials

<table>
<thead>
<tr>
<th>Learning Objectives</th>
<th>Learning Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recognize the impact of the fake news on daily life</td>
<td>Course introduction video, discussion on daily fake news</td>
</tr>
<tr>
<td>Remember the definition and main features of fake news</td>
<td>Video lecture regarding definition and features, dialogue activity</td>
</tr>
<tr>
<td>Figure out how to identify fake news by a four-step method</td>
<td>A true or false game on fake news identification, video lecture regarding a four-step method</td>
</tr>
<tr>
<td>Apply the four-step method to identify real news from cases</td>
<td>Case study</td>
</tr>
</tbody>
</table>

Table 2. Design Examples on Social Presence, Cognitive Presence, and Teaching Presence

<table>
<thead>
<tr>
<th>Element</th>
<th>Categories</th>
<th>Design example</th>
<th>Chatbot prompts example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social presence</td>
<td>Personal/affective</td>
<td>Use emoji</td>
<td>“I’m your learning partner 😊.”</td>
</tr>
<tr>
<td></td>
<td>Open communication</td>
<td>Use vocatives</td>
<td>“Hi Irene, the next video is about the feature of fake news.”</td>
</tr>
<tr>
<td></td>
<td>Group cohesion</td>
<td>Agreements when user gives right answer</td>
<td>“I have the same idea with you.”</td>
</tr>
<tr>
<td>Cognitive presence</td>
<td>Triggering event</td>
<td>Discussion about a fake news example</td>
<td>“Do you think this news is fake or real?”</td>
</tr>
<tr>
<td></td>
<td>Exploration</td>
<td>Dialogue regarding the reason of fake news</td>
<td>“Why do you think it is a true news?”</td>
</tr>
<tr>
<td></td>
<td>Integration</td>
<td>Quizzes to recall the features of fake news</td>
<td>“Can you tell me one of the features of fake news?”</td>
</tr>
<tr>
<td></td>
<td>Resolution</td>
<td>Use 4-step method</td>
<td>“You may need to check the author of this news.”</td>
</tr>
<tr>
<td>Teaching presence</td>
<td>Design and organization</td>
<td>Illustrate learning outcomes in intro-video</td>
<td>“Please watch the introduction video to know your learning objectives.”</td>
</tr>
<tr>
<td></td>
<td>Facilitating discourse</td>
<td>Provide instruction step-by-step</td>
<td>“Now we’re going to explore a 4-step method.”</td>
</tr>
<tr>
<td></td>
<td>Direct instruction</td>
<td>Watch videos</td>
<td>“Please tell me if you finish watch the video.”</td>
</tr>
</tbody>
</table>

2.2 Development of the MIM learning environment

The development of Fake News bot was supported by three software: Dialogflow, Respond.io (which was named Rocketbot before), and WeChat. Dialogflow is flexible enough to be integrated into most of the mainstream platforms in the form of webpage and mobile applications (Bollweg, Kurzke, Shahriar, & Weber, 2018). The visual development dashboard of Dialogflow helps reduce the challenges for non-programming users to design their chatbots. The monthly active users for WeChat is over 1.2 billion by the second quarter of 2020 (Statista, 2020), which takes the market of instant messaging application in China. Students chatted with a WeChat account for receiving the instructional materials
The logic and sequence of interaction between chatbot and students were structured within Dialogflow (developer), which sent videos, pictures, and information to users. The Respond.io as a database stored the instructional videos and pictures as files and provided the file ID as a parameter to Dialogflow. Respond.io supported a workplace to embed chatbot form Dialogflow to WeChat, and automatically collected the chatbot-student dialogue records.

3. Learning with Fake News Bot

3.1 Instrumentation

We conducted two rounds of implementation and testing with the Fake News bot. The first one is an exploratory case to test the operation of the chatbot. In the second round, a five-point Likert scale survey was collected to investigate students’ perceptions concerning the Fake News bot designed by the CoI framework. In both rounds, participants were expected to interact with Fake News bot without any time limit. After students logged in the WeChat application on their mobile phones, they could find the account of fake new bots and started their learning. In Phase One, 43 participants, including university students and recent graduates, gained access to the Fake News bot in WeChat by following the account of the bot. The consent form was embedded in the chatbot before learning materials were presented to the participants. After the students agreed to participate in the intervention, the learning material and interaction would start. The dialogue records indicated that after receiving the consent form, 37% (16 out of 43) participants stopped exploring the chatbot because they were not engaged with it. The other 27 students continued until the link of an introduction video popped up, and subsequently 48% (13 out of 27) stopped going on further. Nine students stopped interacting because they felt “disengaged”, while four students explained that the bot could not understand their inputs (“do not understand”). The records finally revealed that only 6 participants finished all the learning activities.

Initial findings from Phase One indicated two main challenges for students to continue interaction with the Fake News bot. First, students had to open the pop-up materials (such as the link of videos) to jump to another page. Students found this inconvenient and hence they were not interested in continuing with the lesson. The second challenge is related to the design of chatbot. Chatbot with a limited database cannot understand students’ inputs. Therefore, with the bugs and wrong responses from chatbot, the learning environment was less convincing that students cannot gain a sense of immersion during the learning process.

3.2 Chatbot Revision and Evaluation

According to the findings from the exploratory intervention, several actions were taken to improve the Fake News bot. First, the pop-up link of intro-video was replaced by an embedded video to decrease the transferring between different learning interfaces and, therefore, to continue the student-chatbot interaction. Second, hints were added to increase the accuracy of student-chatbot interaction. For instance, once students cannot answer the questions or students gave unpredictable answers, the chatbot would send hints to lead students to work out the tasks.

During the second round, 40 volunteers participated in the conversation with chatbot. 28 participants undertook the online Five-point Likert scale survey, which was presented immediately after they finished the online learning with the Fake News bot. Since the original 34-item CoI measurement (Arbaugh et al., 2008, p.135) focused on the online learning community rather than a specific learning activity, we revised the questionnaire to a 26-item one to analyse the effect of three components of CoI framework on Fake News bot design. In this study, the Cronbach alpha coefficient for the revised 26-items survey showed that cognitive presence was 0.984, and social presence was 0.906, and teaching presence was 0.951, which were consistent with the original measurement.

3.3 Results of Students’ perceptions of the three CoI components

Teaching presence was measured by 9 items, such as “the bot clearly communicated important course topics”, and “the bot helped to keep course participants engaged and participating in productive dialogue”. Social presence was measured by 6 items, for example including “I felt comfortable
disagreeing with the bot while still maintaining a sense of trust” and “I felt that my point of view was acknowledged by the bot”. Cognitive presence was measured with 11 items, such as “reflection on course content and discussions helped me understand fundamental concepts in this class.” Social presence mainly existed between the chatbot and the user in the MIM chatbot based learning environment. The highest mean (M = 4.68, SD = 0.548) among teaching presence items indicated that chatbot was perceived to provide good organization and guidance. Students appreciated the use of emojis during the conversation (M = 4.46, SD = 0.838). Students perceived the reflection on course content and discussion helped them understand the fake news concepts (M = 4.54, SD = 0.576).

4. Lesson Learnt from the Fake News Bot

This study explores the development of an educational chatbot underpinned by the CoI framework. With two-round evaluation of students’ learning with the Fake News bot, we conclude several actions and strategies to implement chatbot into educational environment effectively.

The first principle refers to training chatbot to be a smart partner by increasing the accuracy of chatbot responses. A training phase should be included before bringing the educational bot into learning settings. Pilot study with chatbot testing can be conducted simultaneously to collect students’ real learning records as unpredictable information to enlarge the database of chatbot. The second principle is to increase students’ engagement by adding students’ ownership during interaction and structuring learning materials in one interface (e.g., install video in chatbot). For example, when chatbot questions students with real-life cases, several hints can be delivered to students and give them more choices to answer. For instance, the Fake News bot can provide daily news in differing areas (e.g., sports, nature, political issues), and students can choose their interested category. Likewise, teachers can combine all learning materials in one interface (i.e., chatbot) to diminish the frustrating transfer during the learning process. Third, educators can use the CoI framework as a practical theoretical perspective to design educational chatbots in online learning environment. The social presence can be embedded with the use of emojis throughout student-chatbot interaction to enhance the characteristics of chatbot partner. A peer sharing forum can be set up for students’ collaboration. Based on four categories of cognitive presence (triggering event, exploration, integration, and resolution), diverse scaffolding clues can be used. For instance, before instructing the features of fake news, chatbot can discuss a daily news with students as a warm-up activity, which is the triggering event. The teaching presence can be revealed through the clear illustration of learning outcomes, which can set up students’ prediction on the role of chatbot and continue active learning with chatbot thereof.

5. Discussion and Conclusion

This study provides a development of an educational MIM chatbot, Fake News bot, as an online learning partner to instruct students the identification of daily fake news. Buttressed by the Community of Inquiry framework, the Fake News bot was designed to engage students’ learning in a mobile instant messaging learning context. Students’ use of chatbot and their perspective on the instructional design of chatbot were investigated using a two-round evaluation. Although the number of participants and specific learning domain limits the generalization of the results, the findings are encouraging. The Fake News bot revealed its advantages in facilitating students’ knowledge construction. The use of the CoI framework in chatbot provided students a well-organized learning experience. The demonstration of chatbot development in that study is expected to encourage more instructors to design educational chatbots without the requirement of programming skills.

There are several limitations in this study. First, the format of learning materials was limited by the capability of system. For instance, the video should be less than 20M in size and was compressed by WeChat system automatically, which fell users’ learning experience. Second, participants in two rounds were not same and the chatbot in the second round was more advanced, which impacts on students’ perceptions. The other different individual variables were ignored (e.g., students’ chatbot learning experience before and expectation on chatbot). Third, the short intervention period may lessen the effectiveness of the CoI measurement in this study. For future research, we call for more experimental designs in diverse learning domains and contexts to verify the chatbot design with the Community of Inquiry framework.
References


Learning by Problem-Posing as Kit-Building for Structure Understanding of Polynomial Factorization

Sho YAMAMOTO\textsuperscript{a}, Hiroyoshi ENOMOTO\textsuperscript{b}, Yusuke HAYASHI\textsuperscript{c} & Tsukasa HIRASHIMA\textsuperscript{c}

\textsuperscript{a}Faculty of Engineering, Kindai University, Japan
\textsuperscript{b}Graduate School of Engineering, Hiroshima University, Japan
\textsuperscript{c}Graduate School of Advanced Science and Engineering, Hiroshima University, Japan

*yamamoto@hiro.kindai.ac.jp

Abstract: Polynomial factorization is a basic skill required of middle-school students. Students typically learn the skill through solving polynomial factorization problems. In this paper, as a new method to learn the skill, we propose learning by problem-posing factorable polynomial expressions. In this problem-posing, we use the kit-building method, where a student is provided with a set of components such that a structure can be constructed using the components. To learn polynomial factorization, it is crucial to understand the polynomial expression as a structure. The effectiveness of structure understanding in the kit-building method is confirmed in several learning domains, for example, arithmetic and mathematic word problems. In the exercise performed in this study, a student is required to create a polynomial expression that can be factorized using a specified factorization rule, which we call “problem-posing.” Students and teachers accepted that this is as a better form of learning factorization, i.e., through practical use at universities and middle schools.

Keywords: problem-posing, polynomial factorization, learning environment, kit-building method

1. Introduction

In this paper, we present a learning environment in which a learner learns factoring polynomial by problem-posing. Two kind of case studies are described: "the practical use for university students" and "the case study for middle school students". The focus of this study is the polynomial factorization in middle schools. The factorization used in this study involves factoring a polynomial with factors such as \(x^2 + 5x + 6 = (x + 2) \cdot (x + 3)\). Therefore, learning factorization means understanding the conditions for factoring such a polynomial or expanding the factorized expression into a polynomial. In general, polynomial factorization is learned using problem-solving exercises. To perform effective problem-solving exercises, several learning environments have been proposed (Kim & Glass, 2004; AbuElouf & Abu-Naser, 2017).

Problem-posing has been suggested as an effective learning method for understanding the solution instead of problem-solving (Polya, 1945; Silver & Cai, 1996). This practice has been adopted in various learning environments, and its variations have been developed as well (Yu et al., 2018; Chang et al., 2012; Nakano et al., 1999). We are continuously investigating a learning environment for problem-posing based on problem structures (Hirashima & Hayashi, 2016). In our learning environments, the structure of the problem is defined, and the decomposed one is assigned to the learner as a kit. Learners can deepen their understanding of the problem structure by building the kits (components) through trial and error while receiving feedback from the learning environment. We call this learning method the kit-build method. In this study, we applied a kit-building method for problem-posing of polynomial factorization and developed a learning environment. This described in Section 2. We also reported the results of two case studies in Section 3 and 4. Section 5 is conclusion.
2. Suggested Problem-posing of Polynomial Factorization

2.1 Difficulty of Problem-posing of Polynomial Factorization

Regarding polynomial factorization, problem-posing is a promising learning activity for the same reason, but it is not easy to realize it. The reason is that in the factorization if a student first considers the factorized answer, the student is able to expand it to create a problem. The expansion is usually much easier than the factorization. Practical research of problem-posing of polynomial factorization in a classroom reported that students were able to make several complex problems of polynomial factorization which seemed too difficult for the students to solve (Okiyama, 2011). The result suggests that the students often posed problems by expanding answers. The research concluded that it was not easy to conduct problem-posing of polynomial factorization as a learning activity.

We have continued to develop an original learning environment by problem-posing of the kit-build method. The effectiveness of structure understanding using the kit-building method has been confirmed in several learning domains, e.g., arithmetic and mathematic word problems (Yamamoto & Hirashima, 2017). Therefore, we propose the factorization question learning based on the kit-building method as an effective learning method.

2.2 Suggested Problem-posing of Polynomial Factorization and its Learning Environment

Table 1 shows each step of the suggested exercise. In Step 1 of this exercise, the learner performs the factorization using the formula. This step is intended to review the formula. Figure 1 (a) is shown an interface of exercise for Step 1. The learner inputs and deletes an answer by tapping the blank box to the right of “Equal” and pressing each input button. If the learner taps the diagnosis button in Steps 1, two type of feedbacks are returned: “correct answer” and “calculation error.”

In Step 2, the learner must change the assigned polynomial to a polynomial to which the proposed factorization formula can be applied. The learner is allowed to change the coefficients of the polynomial. A set of changeable coefficients is provided to the learners. As this polynomial is a factorization problem, we call this exercise “problem-posing.” In this activity, the learner must be aware of the conditions under which the assigned factorization formula is applicable. In Step 2, if the learner changes one coefficient appropriately, he/she will obtain the correct answer. In this step, the learner is required to factor the changed polynomial in addition to changing the polynomial. In the step 3 exercise, the learner needs to change some coefficients.

Figure 1 (b) is shown an interface of exercise for Step 2, 3. In this interface, an input area for changing the assigned polynomial is added below the assigned polynomial. When the learner taps the coefficient in this area, a number that can be selected is presented in a pull-down format. By selecting a specific number, the learner can change the assigned polynomial. After changing the assigned polynomial such that it can be factored, the learner inputs the factorized formula similarly as in Step 1. If the learner touches the diagnosis button in Steps 2 and 3, three feedbacks are returned: “correct answer,” “calculation error,” and “polynomial change error.”

This exercise lets learners change non-factorable problems to factorable problems and factor the changed problem for verifying. Through these change activities, the student is promoted to consider the conditions of application of the solution based on the structure of polynomial expressions. To let a student focus on the differences between the original problem and a posed problem, the student is provided components for the change and requested to pose a new problem by using the components. Therefore, they cannot pose a proper problem without explicitly being aware of the factors through the operation of the component. By performing this exercise in each formula, the learner will likely

<table>
<thead>
<tr>
<th>Step 1</th>
<th>Polynomial Factorization</th>
<th>Example</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>3x+3y = ?</td>
<td>3(x+y)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Step 2</th>
<th>Change Polynomial (one coefficient) &amp; Polynomial Factorization (select 4 or 5)</th>
<th>4x+5y = ?</th>
<th>4x+4y</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(select from 2 to 9)</td>
<td>= 4(x+y)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Step 3</th>
<th>Change Polynomial (some coefficient) &amp; Polynomial Factorization (select from 2 to 9)</th>
<th>4x+5y = ?</th>
<th>7x+7y</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(select from 2 to 9)</td>
<td>= 7(x+y)</td>
<td></td>
</tr>
</tbody>
</table>
be cognizant of the conditions for applying each factorization formula. This study targets middle school students. Therefore, five formulas used in the learning environment are as follows: \(ax + ay = a(x+y)\); \(x^2 + (a + b) x + a^2b = (x + a)(x + b)\); \(x^2 - a^2 = (x + a)(x - a)\); \(x^2 + 2a*x + a^2 = (x + a)^2\); and \(x^2 - 2a*x + a^2 = (x - a)^2\). These five formulas were implemented in the learning environment as Levels 1-5. The students learn at each level through the exercises in Steps 1-3.

![Interface of Exercise in Step1](image1.png)

![Interface of Exercise in Step2 and 3](image2.png)

(a) Interface of Exercise in Step1

(b) Interface of Exercise in Step2 and 3

Figure 1. Interface of exercise in learning environment.

3. Experimental Use at University

3.1 Procedure

The participants were 17 undergraduate and graduate students in the engineering department of the university. We would like to verify that the suggested exercise is useful for learners who have already learned factorization. The experiment was a pre-post method, and tests were performed before and after using the learning environment. Participants answered the questionnaire after the test. The times for each activity are as follows: The pretest and post-test required 19 min; the use of the learning environment required 20 min; and providing answers to the questionnaire required 5 min.

In this exercise, participants were required to answer the tasks of levels 1 and 2 only owing to time limitations. If participants had time remaining to use the learning environment, we instructed them to learn repeatedly through assignments from levels 1 and 2.

3.2 Pre-test and Post-test

We conducted three pre-tests and post-tests. The first one is a factorization problem solving test. The easiest problem is of the same difficulty as the problem implemented in the learning environment. Other problems are more difficult than those implemented in the learning environment. These kinds of problem are used in high school learning. The second test is the problem-posing of the factorizing polynomial test. This test is the same type as the exercise performed by the learner in the learning environment. The participant was assigned two polynomials. The learner was required to change each assigned polynomial to pose four factorizable problems in each. The participant was allowed to change only one coefficient when posing a polynomial that can be factorized. The third one is a factorization problem-posing test from scratch. The participants were required to pose four problems.

3.3 Results

First, we present the log analysis of the learning environment. The system log was obtained from 17 participants. From the analysis result, it was discovered that all 17 participants completed the level 1 and 2 exercises. The average number of correct answers and error were 8.65 and 3.94.

Second, the factorization problem-solving test did not differ significantly different between the pre- and post-tests (Wilcoxon signed-rank test, \(p = .343 > .05\)). The percentage of correct answers increased only slightly from 48% to 54% overall. However, the correct answer rate was 90% for the simplest problems that were implemented in the learning environment. The percentage of correct
answers increased only slightly from 47.5% to 53.8% in high difficulty problems. This result showed that the participants successfully solved the difficult factorization.

Third, we show the results of problem-posing in the factoring polynomial test and the factorization problem-posing test (Table 2). Here, we investigated the number of posed problems that can be solved by a formula other than the one used in the exercise (called advanced problem). The number of this advanced posed problems is significantly different between the pre- and post-tests (Wilcoxon signed-rank test, $p = .004 < .05$). In the factorization problem-posing test, a significant difference was observed in the number of problems posed between the pre- and post-tests (Wilcoxon signed-rank test, $p = .011 < .05$). Furthermore, the number of posed advanced problem between the pre- and post-tests differed significantly (Wilcoxon signed-rank test, $p = .005 < .05$). These results showed that the participants successfully posed the difficult factorization problem.

Finally, the questionnaire and its results are shown in Figure 2. This questionnaire was implemented based on a four-point scale (strongly agree, almost agree, almost disagree, and strongly disagree). While the feedback was not agreed highly, the usefulness of the learning environment for factorization learning was highly evaluated.

From the results of this practical use, we concluded the following: (a) the suggested exercise was useful for learning factorization for university students who have already learned factorization; (b) the suggested exercise may promote the learner to awareness for transfer. This result suggests that learners can deepen their understanding of the factorization formula by our leaning environment.

<table>
<thead>
<tr>
<th>Table 2. The Score of Each Problem-posing Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test type</td>
</tr>
<tr>
<td>Problem-posing in the factoring polynomial test</td>
</tr>
<tr>
<td>Factorization problem-posing test</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

*p < .05

Figure 2. The contents and answers of the questionnaire.

4. Case Study at Middle School

4.1 Procedure

After conducting the experimental use, we explained our learning environment to middle school teachers. We decided to conduct a case study based on 37 third-grade middle-school students, who were the better problem solvers among all third graders in a particular school. Furthermore, they have already learned polynomial factorization. The purpose of this case study is to verify that the target learner can learn factorization in the proposed learning environment.

In this case study, we first explained the method to use the system and provided an Android tablet within approximately 10 min. Next, the students learned polynomial factorization in our learning environment for approximately 30 min; finally, they answered the questionnaire in 5 min. This exercise comprised only 15 problems with five levels. The learners who completed the answer were asked to use
this system again or to use another system. As using another system does not affect the analysis of this case study, we will omit the explanation of the other system used. In addition, when students decided to use our learning environment again, they were instructed to consider another answer.

4.2 Time of exercise

We were able to collect the appropriate logs for 36 students. The average time taken for this exercise was 23 min 33 s, including the time for a second learning. Figure 3 shows the distribution of the learner’s practice time in our learning environment. The usage time was polarized significantly, and we speculated that the quality of the exercises differed by group. Therefore, 13 students who had been learning beyond the average usage time were grouped as the long-time usage group (L-group), whereas 23 students who required less than the average usage time were grouped as the short-time usage group (S-group). In addition, by investigating the number of diagnoses performed every 10 min in each group, we discovered that approximately 10 - 14 times of diagnosis was performed in each term. This fact suggested that the student has learned continuously in this case study.

![Figure 3. Learning Environment Time and Number of Participants in Each Time.](image)

4.3 Result of Number of Correctness and Incorrectness on Learning Environment

The percentage of correct answers in all exercises using our learning environment were 72.2% and 87.3% in the L-group and S-group, respectively. A significant difference was observed between these percentage of correct answers rates (Wilcoxon signed-rank test, \( p = .0002 < .01 \)). In addition, the percentage of correct answers in the first and second trials of each group was analyzed, in which those of the L-group were 71.1% and 79%, respectively. Meanwhile, those of the S-group were 88.5% and 87.0%, respectively. No significant difference was observed between the first and second correct accuracy rates in each group. The correct answer rates in first trial are shown in the order of Step 1, Step 2, and Step 3 of each level, where those of the L-group were 93.1%, 80.1%, and 83.8%, respectively; and 98.3%, 92.2%, and 91.6%, respectively, for the S-group.

Finally, we describe the number of errors generated by the learner in our learning environment. A total of 143 errors were discovered in all the students’ exercises. The details are shown in Table 3. Three types of errors occurred: “error of method for changing polynomial,” “error of factoring polynomial,” and “error of changing polynomial.” The first one is the error where the polynomial is not changed as shown in the assignment. For example, the assignment requires changing two coefficients, but the learner does not change any of them. The second one is an error in polynomial factorization. The third one is an error where the polynomial change is wrong. This error includes cases where the learner has posed a problem that cannot be factored or that cannot be solved with the given formula.

<table>
<thead>
<tr>
<th>Table 3. Total Number of Each Errors in Each Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error of method for changing polynomial</td>
</tr>
<tr>
<td>Error of factoring polynomial</td>
</tr>
<tr>
<td>Error of changing polynomial</td>
</tr>
<tr>
<td>Step 1</td>
</tr>
<tr>
<td>Long-term use</td>
</tr>
<tr>
<td>Short-term use</td>
</tr>
</tbody>
</table>

4.4 Questionnaire Results

Finally, the questionnaire results are shown in Figure 4. The contents of questionnaire are the same as that shown in Figure 2. For almost all questions, the S-group and L-group answered positively.
Immediately after the class, we sought the opinions of two teachers, i.e., a mathematics teacher and a class teacher, regarding the practical use of the exercises. They reported that the exercises were meaningful for understanding factorization, and that such exercises would be impossible without our learning environment. Moreover, the teachers stated that assigning one lecture period to solve the exercises of this system would be worthwhile.

Compared with the performance of the S-group, that of the L-group was lower. The L-group had more errors than the S-group, and the exercise time was longer. However, the S-group more appreciates our learning environment than the L-group. This result suggests that our learning environment may be more useful in a slow learner.

From these results, it was confirmed that both the learners and teachers recognized that the exercises of the proposed learning environment were useful for understanding factorization.

![Figure 4. Results of questionnaire.](image)

(a) Answers for Long-term Use Group  
(b) Answers for Short-term Use Group

5. Conclusions and future works

In this paper, we proposed a learning environment for problem-posing a polynomial factorization by kit-build method. Furthermore, we reported the results of practical use for university students and of case studies for middle school students. The results of these case studies have been suggested that our learning environment may be effective in understanding mechanism of factorization.

For future studies, we are considering improving the feedback of learning environment. Furthermore, we plan to conduct a case study based on an experimental group and a control group and a case study targeted a slow learner for verifying learning effect of our learning environment.

References

Predicting end-of-session actions considering the information of learning materials

Daichi Takehara*
* Aidemy Inc., Japan
* takehara-d@aidemy.co.jp

Abstract: To provide learners with a better learning experience in online educational systems, it is meaningful to understand and model learners’ actions. The actions of learners ending their learning sessions and leaving the systems, which we denote as the end-of-session actions, is important to understand. Modeling the end-of-session actions can lead to useful applications, such as optimizing the way learning materials are presented and interventions that can appropriately help learners. This paper addresses the problem of predicting end-of-session actions in online educational systems. While previous studies have mainly focused on the learners’ behavior in the systems, this paper focuses on incorporating the information of learning materials into the prediction model. Learning material features were extracted by considering multiple perspectives in the learning materials, including their order in the course and their texts. The experiment was conducted using actual user log data from the programming learning system. The experiment demonstrates the effectiveness of incorporating learning material features into the prediction models and analyzed their contribution to the prediction accuracy.

Keywords: end-of-session action prediction, learning materials modeling, feature extraction, machine learning

1. Introduction

Recently, online educational systems, such as Massive Open Online Courses (MOOCs), have been established and widely used. In many cases, online educational systems allow learners to actively choose the time to study and the order of learning materials to be taken, depending on their learning style. Learners voluntarily access the system to begin and end their learning. Given that learners use the system in this way, it is worthwhile to model the actions of learners ending their learning sessions and leaving the systems, which we denote as end-of-session actions. By being able to model these accurately, systems can provide a better learning experience for learners through optimizing the way learning materials are presented and interventions that can appropriately help learners.

Several studies have addressed the problem of modeling and predicting end-of-session actions of learners in online educational systems (Karumbaiah et al., 2018; Hansen et al., 2019). Karumbaiah et al. (2018) proposed a method that predicted the end of students playing a learning game called Physics Playground developed for physics education for middle school students. The authors designed 101 features, focusing on progress and experiences based on learners’ behavioral history in the game. The authors used the obtained features in a prediction model with a gradient-boosting tree. Hansen et al. (2019) presented the problem of determining when a session ends by modeling the probability that an action will be the end of the session. Through modeling by long short-term memory (LSTM) (Hochreiter et al., 1997), the authors showed the effectiveness of considering learners’ long-term behavioral history in the problem. Although these previous studies paid attention to learners’ behavior in the educational systems, insufficient consideration was given to incorporating the information of the learning materials into the model.

This paper addresses the prediction of end-of-session actions incorporating the information of learning materials. This paper proposes a model that predicts the probability if the learner’s answer to an exercise will be the end-of-session action. The proposed method extracts not only the features related to the learners’ actions, i.e., the learner features, but also the features related to the learning materials, i.e., the learning material features, from the log data stored in the system. The learning material features
are extracted by considering multiple perspectives in the learning materials, including their order in the course and their texts. Based on the extracted features, the proposed method learns the prediction model by machine learning, linear regression (Seber et al., 2012), random forest (Breiman, 2001), and gradient-boosting tree (Friedman, 2001) algorithms. The experiment was conducted using actual learners’ log data on Aidemy, an online programming learning system (the system is described in detail in section 2.1). The experiment verified the effectiveness of incorporating the learning material features into the prediction of end-of-session actions and analyze their contribution to the prediction accuracy of them in detail.

2. Data

2.1 Aidemy

In this paper, we used the log data of learners on Aidemy. Aidemy is an online programming learning system that allows learners to acquire skills and knowledge about data science, such as statistics and machine learning. The main feature is that learners can learn not only by reading learning materials, but also by coding in the browser editor, as shown in Figure 1. Each course includes a technical topic (e.g., “Introduction to Machine Learning,” “Introduction to Python,” and “Fundamentals of Deep Learning”) and consists of multiple exercises. The types of exercises include multiple choice questions, coding questions, and videos. After registering for the system, learners can purchase and start taking courses that interest them from the 46 courses published as of December 20, 2019.

![Figure 1. The screen where a learner is taking the exercise in Aidemy. The left side of the screen shows the explanation of the exercises, and the right side of the screen shows questions that are answered by inputting the code or the answer to the multiple questions.](image)

2.2 Log Data

The log data used in the experiment is generated when a learner answers an exercise on Aidemy. The log data includes the timestamp, the learner id, the course id, and the exercise id. The course id and the exercise id are associated with information about each learning material, including the title and description of the learning material, the type of exercise, and the Python libraries used. In this paper, the
The proposed method extracts learner features and learning material features from these log data and uses them for training a predictive model.

We define end-of-session actions by the length of time spent on the answer based on the timestamp of the log data. If the length of the answer is more than 15 minutes, the action is considered to be the end-of-session action. Since the length of time spent on the answer of a single exercise on Aidemy often ranges from a few tens of seconds to a few minutes and rarely exceeds 15 minutes, we empirically set the threshold value to 15 minutes.

In the experiment, we collected the log data for the period from January 1, 2019, to November 30, 2019. Learners who had purchased one or more paid courses and completed at least ten exercises were used for the experiment. The collected log data included 1314 learners, 326890 answers, and 35033 end-of-session actions.

3. Predicting End-of-session Actions Considering the Information of Learning Materials

3.1 Problem Setting

This paper addresses the problem of predicting end-of-session actions, as in the paper by Hansen et al. (2019). In this problem, the prediction model outputs the probability if the learner’s answer to the exercise will be an end-of-session action. The training dataset used pairs of a feature extracted from the learner’s answer and a binary label.

3.2 Feature Extraction

The features extracted from the log data are described in Table 1. The features are divided into multiple separate feature groups, learner features, and four learning material features (basic, order, text, and library). Learning material features are divided into subgroups based on their characteristics for more detailed analysis in the experiment.

Table 1. Overview of Features

<table>
<thead>
<tr>
<th>Group</th>
<th>Feature name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learner</td>
<td>learner id</td>
<td>learner identifier</td>
</tr>
<tr>
<td></td>
<td>session num</td>
<td>number of sessions</td>
</tr>
<tr>
<td></td>
<td>stay time in last answer</td>
<td>answering time in the previous answer</td>
</tr>
<tr>
<td></td>
<td>stay time from last session</td>
<td>total answering time since the previous session</td>
</tr>
<tr>
<td></td>
<td>session length</td>
<td>number of answers in the current session</td>
</tr>
<tr>
<td></td>
<td>mean session length</td>
<td>mean of the number of answers in the learner’s session</td>
</tr>
<tr>
<td></td>
<td>std session length</td>
<td>standard deviation of the number of answers in the learner’s session</td>
</tr>
<tr>
<td></td>
<td>answer time zone</td>
<td>answer time zone (6 different time zones of the day, 4 hours each for the 24 hours in a day)</td>
</tr>
<tr>
<td></td>
<td>learner time zone</td>
<td>ratio of answer time zone of the learner</td>
</tr>
<tr>
<td>Learning material (basic)</td>
<td>course id</td>
<td>course identifier</td>
</tr>
<tr>
<td></td>
<td>exercise id</td>
<td>exercise identifier</td>
</tr>
<tr>
<td></td>
<td>exercise type</td>
<td>types of exercises (choice question, code question, video)</td>
</tr>
<tr>
<td></td>
<td>change exercise</td>
<td>indicator corresponding to whether the exercise is different from the previous one or not</td>
</tr>
<tr>
<td></td>
<td>change course</td>
<td>indicator corresponding to whether the course is different from the previous one or not</td>
</tr>
<tr>
<td>Learning material (order)</td>
<td>order</td>
<td>order of the exercise (counting from the beginning of the course)</td>
</tr>
<tr>
<td></td>
<td>order from chapter end</td>
<td>order of the exercise (counting from the end of the chapter)</td>
</tr>
<tr>
<td>order from end</td>
<td>order of the exercise (counting from the end of the course)</td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>-------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Learning material (text)</strong></td>
<td>Dense features (5 dimensions), which are extracted by applying the term frequency-inverse document frequency (TFIDF) (Manning et al., 2008) to course and exercise titles and descriptions and and dimensionality reduction by singular value decomposition (SVD) (Halko et al., 2009)</td>
<td></td>
</tr>
<tr>
<td><strong>Learning material (library)</strong></td>
<td>Categorical features of which Python library is being used in each course. It targets the libraries that are being imported by the code input exercises. For example, they include NumPy, pandas, and scikit-learn, etc.</td>
<td></td>
</tr>
</tbody>
</table>

The learner features are extracted based on the history of learners on the systems, which also represent the learners’ behavior and recent attendance.

Learning material features are extracted by considering multiple perspectives of the learning materials. Learning material (basic) contains basic information, such as id and types of learning materials. Learning material (order) contains the information about the order of the exercises in the course. It may contain some meaning depending on the order of the exercises, e.g., exercises at the end of a course or the end of a chapter are likely to be the timing of a learning break. Learning material (text) is extracted by using the text contained in the learning materials and represents their content. Learning materials (libraries) represent the Python libraries used in the course. This information is essential for learners of the programming learning system.

### 3.3 Models

Based on the extracted features, the proposed method learns the prediction model by machine learning algorithms. To learn the prediction model, we used three algorithms: linear regression (Seber et al., 2012), random forest (Breiman, 2001), and gradient-boosting tree (Friedman, 2001) algorithms in the experiment. Linear regression is a simple algorithm that assumes a linear relationship between features and targets. Decision tree-based algorithms, such as random forest and gradient-boosting trees, are advanced algorithms that can effectively learn features. Moreover, a detailed analysis using the importance of the features with Gini importance (Breiman, 2001) can be performed. This analysis is consistent with the primary purpose of this paper, which is to verify the effectiveness of each learning material feature.

### 4. Experiments

In the experiment, we verified the effectiveness of incorporating the information of learning materials into the prediction of end-of-session actions by using actual log data in Aidemy.

#### 4.1 Experimental setup

The evaluation was performed using the time-series cross-validation approach (Hyndman, 2019), which divided the log data into multiple time series. Five datasets were created in the experiment. Each dataset consisted of a 6-month training dataset and a test dataset for the following month (e.g., if the training dataset was collected between January and June 2019, then the test dataset was collected in July 2019.). The evaluation scores were then calculated by averaging the score for each test dataset. Validating the model over multiple periods in this way increases the reliability of the evaluated scores. Five-month test datasets for July, August, September, October, and November 2019, were used in the experiment. The evaluation metric was AUC, which is the area under the receiver operating characteristic (ROC) curve.

The prediction model was trained using multiple combinations of the features described in Section 3.2. Specifically, we compared the use of only the learner features (Learner), the use of one of the learning material features in addition to the learner features (Learner w/ learning material feature group), and the use of all the features (All features). The prediction models were trained with the linear
regression, random forest, and gradient-boosting tree algorithms, as described in 3.3. These algorithms use an implementation of the Python library, Scikit-learn (Pedregosa et al., 2011), and the hyperparameters are used as their default values.

4.2 Results

The experimental results are shown in Table 2. The AUCs of the cases when incorporating learning material features (All features, Learner w/ learning material feature group) were higher than cases when using only learner features (Learner). As a result, the effectiveness of incorporating the learning material features in predicting end-of-session actions was confirmed.

Table 2. Evaluation Results.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Feature*</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Regression</td>
<td>Learner</td>
<td>0.6250</td>
</tr>
<tr>
<td></td>
<td>Learner w/ Learning Material (basic)</td>
<td>0.6369</td>
</tr>
<tr>
<td></td>
<td>Learner w/ Learning Material (order)</td>
<td>0.6251</td>
</tr>
<tr>
<td></td>
<td>Learner w/ Learning Material (text)</td>
<td>0.6540</td>
</tr>
<tr>
<td></td>
<td>Learner w/ Learning Material (library)</td>
<td>0.6393</td>
</tr>
<tr>
<td></td>
<td>All features</td>
<td><strong>0.6651</strong></td>
</tr>
<tr>
<td>Random Forest</td>
<td>Learner</td>
<td>0.6078</td>
</tr>
<tr>
<td></td>
<td>Learner w/ Learning Material (basic)</td>
<td>0.6621</td>
</tr>
<tr>
<td></td>
<td>Learner w/ Learning Material (order)</td>
<td>0.6661</td>
</tr>
<tr>
<td></td>
<td>Learner w/ Learning Material (text)</td>
<td>0.6725</td>
</tr>
<tr>
<td></td>
<td>Learner w/ Learning Material (library)</td>
<td>0.6137</td>
</tr>
<tr>
<td></td>
<td>All features</td>
<td><strong>0.6840</strong></td>
</tr>
<tr>
<td>Gradient-boosting tree</td>
<td>Learner</td>
<td>0.6650</td>
</tr>
<tr>
<td></td>
<td>Learner w/ Learning Material (basic)</td>
<td>0.6928</td>
</tr>
<tr>
<td></td>
<td>Learner w/ Learning Material (order)</td>
<td>0.7067</td>
</tr>
<tr>
<td></td>
<td>Learner w/ Learning Material (text)</td>
<td>0.7073</td>
</tr>
<tr>
<td></td>
<td>Learner w/ Learning Material (library)</td>
<td>0.6697</td>
</tr>
<tr>
<td></td>
<td>All features</td>
<td><strong>0.7155</strong></td>
</tr>
</tbody>
</table>

*The notation of the feature columns corresponds to the group columns in Table 1.

4.3 Further Analysis

The experimental results show that, although all learning material features contribute to improve the prediction accuracy, the degree of their contribution is different. To confirm how each feature works in more detail, we observed the Gini importance (Breiman, 2001) in the gradient-boosting tree as the feature importance. Figure 2 shows the feature importance aggregated by the sum of each feature group shown in Table 1 (left figure) and the top ten features of the higher feature importance (right figure).

Figure 2. Feature importance (left: the feature importance aggregated by the sum of each feature group, right: the top ten features of the higher feature importance). The notation of the vertical axis corresponds to the group and feature columns in Table 1.
Similarly to the degree of improvement in AUCs in Table 2, the importance of Learning material (order) and Learning material (text) was confirmed to be high. Learning material (order) is consistent with the intuitive understanding that delimitations such as the end of the course and the end of the chapter in a course are strongly related to whether learners continue learning or not. In addition, Learning material (text) is a feature that represents semantics about the content of the learning material, such as a technical topic. Therefore, it is suggested that the content of the learning material had a moderate effect on the learner’s end-of-session actions.

5. Conclusions

This paper addressed the problem of predicting end-of-session actions incorporating the information of learning materials. The proposed method extracted not only the learner features, but also the learning material features from the log data stored in the system. Based on the extracted features, the proposed method learned the prediction model by machine learning algorithms.

In the experiment using actual log data in Ayden, the effectiveness of incorporating the learning material features into the prediction of end-of-session actions was confirmed. In addition, the features’ importance and contribution to the predictive model was verified. In this experiment, we can confirm that features related to the order of the exercises in the course and text contained in the learning materials are significant contributors to prediction accuracy.

References

Using Augmented Reality (AR) in Innovating Pedagogy: Students and Psychologists’ Perspectives

Hazel TRAPEROa,e, Ryan EBARDOb, Jypzie CATEDRILLAc, Laiza LIMPINC, Josephine DE LA CUENSta, Cecilia LEAÑOe, Michelle Renee CHINGe

aUniversity of the Philippines Cebu, Philippines
bJosé Rizal University, Philippines
cMindanao State University – General Santos, Philippines
tAsia Pacific College, Philippines
eDe La Salle University Manila, Philippines
*hazel_trapero@dlsu.edu.ph

Abstract: Augmented Reality (AR) is one of the emerging technologies of interest and is seen to be applicable in educational context since it is capable of leading richer learning experiences. However, many pieces of research focused on the technicalities, efficacy of various systems, usability, and design but less on how social interactions, or its social-psychological aspects, are affected by the technology. Given that the findings of prior research necessitated elaboration in the social-psychological aspects of the ethical issues of AR through an expert evaluation, this study was conducted to investigate the social-psychological impacts of this immersive technology to students. This study was also conducted to evaluate the students’ experiences of using AR technology based on the Psychologists’ perspective. It was found out that AR integration to innovate pedagogy could have a significant impact on the learners and the teachers in educational institutions and universities, but there remain a number of challenges and issues on its use.

Keywords: Augmented Reality, Innovating Pedagogy, Ethical Issues, Social Psychology

1. Introduction

Augmented Reality (AR) is one of the emerging technologies of interest and is seen to have a significant impact on the learners and teachers in educational institutions and universities (Chang, Hsu, & Wu, 2016; Sáez-López, Cózar-Gutierrez, González-Calero, & Carrasco, 2020; EDUCAUSE, 2018). Meta-analysis researches provide evidence of the progress of this emerging and innovative technology in teaching and learning processes (Ozdemir, Sahin, Arcagok, & Demir, 2018; Hantono, Nugroho, & Santosa, 2018). In the learning process, it boosts students’ academic achievement compared with engaging in traditional learning methods (Ozdemir et al., 2018). However, many pieces of research focused on the technicalities, efficacy of various systems, usability, and design but less on how social interactions are affected by the technology or its social-psychological aspects (Miller et al., 2019). Social psychology is defined as the pursuit to understand and explain how the actual, imagined, or implied presence of others influences the thoughts, feelings, and behavior of individuals (Allport, 1979). Given that the findings of prior research (Trapero, 2018) necessitated elaboration in the social-psychological aspects of the ethical issues of AR through an expert evaluation, this study was conducted to investigate the social-psychological impacts of this immersive technology to students.

2. Methodology

In the previous study (Trapero, 2018), thirty-Nine (39) Grade 7 students from the University of the Philippines Cebu (UPC) were identified as the respondents of the study since they are the youngest group among high school students. This is because young individuals now are becoming accustomed to
vast information access (Shamir, 2013) and because the study aimed to evaluate the ethical issues of AR in the early stage of the learning process. The qualitative data from the previous study were collated and analyzed to further elucidate the study. These data were evaluated by two psychologists for evaluation and analysis. Findings, conclusions, and recommendations were then formulated and presented.

3. Results and Findings

3.1 Psychologists’ Social-Psychological Evaluation of Students’ Experiences with AR Apps

P1: “From the psychological point of view, whenever we have a problem, regardless whether we are a student or not, we are taught about how important it is to connect with the support system like family, close friends, professionals, and other people whom we can trust. With the advent of AR technology, engaging in it is a setback to our face-to-face communication, which is supposedly considered as therapeutic. AR can alienate us further. For example, a person does not walk, jog, or talk with friends anymore because he/she is already engrossed and busy with an AR app. A person will not engage in sports anymore, which might lead to negative consequences in health, will not connect with nature like climbing mountains, going to the beach, and other outdoor activities which should supposedly be done to stay healthy and to unwind from stressful situations.”

Using an AR app may lead to a risk of developing a personality disorder; however, this is highly theoretical and needs actual research. Whenever we have a problem, there is always a limit as to how much we can handle it. A series of overwhelming problems, unfortunate events, and others will sometimes force us to dissociate from our “real” problematic self and create an “imaginary” self where all our problems do not exist. It is like our last resort to survive, denying that all our problems existed, like in Dissociative Identity Disorder. AR can be a trigger, or should I say, can encourage people to create an imaginary self since face-to-face communication is already absent.

For instance, an introvert and problematic student with a dysfunctional family system are introduced to AR which may assume an imaginary character who is successful, has no problem, and who possesses qualities that are difficult to achieve in real life. There is a possibility that the like between fantasy and reality will get blurred for that student. This will also lead him/her to start imagining that he/she is that AR character and act like one and deny the existence of his/her self. This is already an alarming personality condition and is not good anymore.

Another risk is the risk to trigger a trauma if a student has a childhood traumatic experience such as molestation, terrible accident, and others, which was not taken away from him/her because of the absence of therapy. The traumatic memories are just suppressed by a person and any significant keyword, may it be an event, image, person/character with the same behavior as a perpetrator (in case of trauma due to rape or molestation) and others that a student will experience or will see in an AR app will immediately trigger the trauma to surface. So, if an AR app is used to facilitate learning in the classroom setting, that is a good intention, but if there are trauma cases that the school or the teacher is unaware of, it will drive the student not to go to school anymore, instead of learning.”

P2: “Augmented Reality is an interesting technological innovation, however, the biggest challenges on the risks in Table 2 would be: 1) they are being prone to accident, 2) an AR app makes them addicted to it, 3) it makes them lazy, and 4) it increases the possibility of acquiring a sickness. So, if I were a parent, I will not allow my child to engage in it, unless I get the trust and assurance for the safety of my child.”

3.2. Psychologists’ Recommended Precautions

P1: - Educate the students about the importance/benefits of face-to-face communication, personal interactions, activities, sports, and others to their psychological, physiological well-being, among others.
- School administrators/teachers should craft strict policies to limit AR app usage of students (e.g.: during classes no using of smartphones unless allowed by the teacher, if they connect to the
internet using the public school Wi-Fi, the school can restrict access to AR apps or impose a time limit in their access to Wi-Fi. Another way is to design group activities, assignments, and other classroom activities that will discourage them from using their smartphones and force them to talk to their classmates, friends, or interview people.

- Parents can impose their own rules as well, outside school premises, it is the responsibility of the parents to monitor and limit smartphone usage or AR apps usage. (e.g.: by 7 pm no more using of smartphones, strictly reading or study for exams).
- School psychologists or counselors can help as well through counseling and other psychological interventions in cases where teachers and parents are no longer effective in limiting the student’s AR exposure.

P2: - There should be a guidance counselor to assure that the students will not get addicted and prevent them from becoming socially isolated individuals.
- Parents should encourage the students to limit the use of AR apps and continue social development to eliminate the issue of the negative changes in behavior.

Conclusion

Augmented Reality integration to innovate pedagogy could have a significant impact on the learners and the teachers in the educational institutions and university, but there remain a number of challenges and issues on its use. It can increase the students’ motivation in the learning process, provide fun and stimulation, enhance their learning interests and self-confidence, and serves as a useful tool for teachers in improving their teaching effectiveness since it enables the integration of the real-world with the learning environment. However, risks and disadvantages on young individuals are also apparent, particularly in social, visual, and motor development. Thus, rules on limiting the AR app usage should be imposed in school and at home, since engaging in it is a setback to their face-to-face communication and getting hooked to it may alienate them further. Parents and teachers are advised to constantly monitor their child’s use of immersive technologies like AR apps. Lastly, psychological intervention should be provided to give assurance that students to prevent addiction, mitigate the risk of developing a personality disorder, and to prevent them from becoming socially isolated individuals.

References

Accuracy-aware Deep Knowledge Tracing with Knowledge State Vector Loss

Qiushi PAN & Taro TEZUKA
University of Tsukuba, Japan
qiugits@klis.tsukuba.ac.jp, tezuka@slis.tsukuba.ac.jp

Abstract: In major e-learning platforms such as intelligent tutoring systems (ITSs) and massive open online courses (MOOCs), the students are often recommended what course materials to take based on their past interactions. Knowledge Tracing (KT) is the task of modeling students’ academic abilities. Given a sequence of student’s learning history, it predicts how well they will perform in the next interaction. Deep Knowledge Tracing (DKT) uses a recurrent neural network (RNN) to capture the underlying structure of the student's understanding. In this paper, we point out the accuracy rate problem that the model won't reproduce the accuracy ratio. This is a limitation of the existing loss function in DKT that it only learns the probability of correctly answering a problem in the next interaction. We introduced the Knowledge State Vector loss, which captures the accuracy rate of all knowledge concepts, to measure and train the model.

Keywords: Deep Knowledge Tracing, Machine Learning, Educational Data Mining

1. Background

On e-learning platforms such as intelligent tutoring systems and massive open online courses, students and teachers can benefit from a more personalized educational system. Knowledge Tracing is a task to model the dynamics of a student’s acquisition of knowledge concepts (KC). KC is a general term that represents a unit of knowledge, and can also be expressed as skill, ability, etc. Given a sequence of a student’s learning history, it predicts how they will perform in the next interaction. Bayesian Knowledge Tracing (BKT) (Corbett et al., 1994) tackles this task with an interpretable model that requires domain specialists to tune its parameters. Deep Knowledge Tracing (DKT) (Piech et al., 2015) uses a recurrent neural network (RNN) (Zaremba et al., 2014). It achieves higher accuracy without a domain specialist, but its parameters remain difficult to be explained. To optimize for the next interaction result, the loss function of a DKT model is

\[ L = \sum_{t=1}^{T} \kappa_y y_t \delta_m(q_{t+1}), a_{t+1} \]

where \( t \) is a time step in the learning history, \( y_t \) is an \( m \)-dimensional vector representing the prediction for each KC’s probability being correctly answered. \( m \) is the total number of KCs. \( q_{t+1} \) is a scalar index representing the label of the KC studied at time \( t + 1 \), and \( \delta_m \) converts the label to an \( m \)-dimensional one-hot vector. \( a_{t+1} \) is the ground truth answer result that indicates whether the student answered correctly (1) or not (0). The loss \( \kappa \) is binary cross-entropy. Input \( x_t \) is given as a one-hot representation such that \( x_t = \delta_{2m}(q_{t+1} + a_m) \). Based on the method of compressed sensing, we embedded \( x_t \in \{0,1\}^{2m} \) into a ceil(log \( 2m \)) length vector and passed to the RNN cells that recursively outputs \( y_t \). Among a number of problems reported in DKT, the reconstruction problem is where the model does not reproduce the last study log in the predictions, and the wavy transition problem is that the predicted performance fluctuates rapidly over time (Yeung et al., 2018).

2. Method
We point out the accuracy rate problem, the inability to trace the accuracy performance of students properly. Although the model with the existing loss function $L$ models the probability of answering the immediately following KC $q^m_{t+1}$ correctly, it does not model the correct answer rate. This comes from the limitation of the existing loss function $L$ that in order to make an estimation, it requires knowing what question the student will choose to solve in the next step. This means, in turn, the model can not predict well about a KC $q^β$ if it does not appear at $t+1$. Also, the probability of answering correctly at $t+1$ is not ideal as the assessment of academic ability since it can easily fluctuate over time. We consider the accuracy rate is more suitable as a representation of students’ academic skills.

To solve this problem, we propose a novel loss function that optimizes the accuracy rate as a regularization problem. It is reasonable that in the first place, the loss function is invented to utilize the next answered KC result because it is the only information we can directly get from the dataset. In this paper, by calculating the average accuracy rate for all KCs, we can access target data of KCs not limited to what is actually answered in the next time step.

Our new loss function is based on the assumption that at each time point, a student has a probability distribution on whether he or she can solve each problem correctly. We call this a knowledge state vector (KSV). The KSV loss is considered more suitable as an optimization target because it considers all information in the given input when optimizing to the next time step.

Let $T$ be a sequence length. $δ_m(q_{t+1})$ represents an $m$-dimensional one-hot representation of the skill involved in solving a problem at timestamp $t$. The sum of $δ_m(q_t)$ over time is a vector of the frequency of the skills that appear in the sequence. Since $a_sδ_m(q_s)$ is the product of a scalar and vector, their sum is the frequency of skills $q_1, ..., q_t$ of correctly solved problems. Let $m$ be the number of KCs, $2$ be the set $\{0, 1\}$, and be the set of integers. The simplest KSV loss is

$$L_{ksv} = \sum_{t=1}^{T} \kappa \left( y_t \circ \sum_{s=2}^{t+1} \delta_m(q_s), \sum_{s=2}^{t+1} a_sδ_m(q_s) \right)$$

where $\delta_m(q_s) \in 2^m$, $\sum_{s=2}^{t+1} δ_m(q_s) \in Z^m$, $a_sδ_m(q_s) \in 2^m$, $\sum_{s=2}^{t+1} a_sδ_m(q_s) \in Z^m$. $\circ$ is the Hadamard product. The original loss function takes skill frequency and accuracy rate as arguments. The loss function $\kappa$ is a mean squared error. By calculating the loss, we can optimize the model such that its prediction will be closer to the accuracy rate. To model the process of forgetting, we introduce an alternative model that has a parameter $β$ that puts exponential weights to questions that were answered recently.

$$L_{ksv, β} = \sum_{t=1}^{T} \kappa \left( y_t \circ \sum_{s=2}^{t+1} β^sδ_m(q_s), \sum_{s=2}^{t+1} a_sβ^sδ_m(q_s) \right)$$

In the experiments, we used $β = 1.2$. Equation 1 is an example of Equation 2, where $β$ is set to 1. Finally, the updated loss function is formulated as $L' = L + λ_{ksv}L_{ksv, β}$ where $L$ is the existing loss. $λ_{ksv}$ is a regularization parameter for KSV loss $L_{ksv}$.

3. Results
Figure 1. Learning curves of the proposed method for different datasets. Left is about the AUC, and right is about the KSV loss. The horizontal axis represents the iteration epoch number.

Table 1. Comparison of DKT results.

<table>
<thead>
<tr>
<th>Assisments 2009</th>
<th>Simulated-5</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\lambda_{\text{ksv}})</td>
<td>(\lambda_{\text{ksv}})</td>
</tr>
<tr>
<td>0.00</td>
<td>0.8010</td>
</tr>
<tr>
<td>0.02</td>
<td>0.7893</td>
</tr>
<tr>
<td>50.00</td>
<td>0.7132</td>
</tr>
</tbody>
</table>

We used the ASSISTments "Skill Builder" Dataset 2009–2010 (Feng et al., 2009). Each problem in this dataset is manually tagged with a single required skill among 124 knowledge concepts. We also used Simulated-5 (Piech et al., 2015), a simulation of 4,000 virtual students solving 50 problems. For comparison, we reconstructed the DKT as well as the waviness indicators \(w_1\) and \(w_2\) and current time step AUC \(C\) (Yeung et al., 2018). The baseline DKT is set up on the basis of previous research. We used a hidden dimension size of 200 and a batch size of 128. The learning rate is set to 0.05. We experimented for \(\lambda_{\text{ksv}} \in \{0, 0.02, 50\}\). Since the AUC is the result of categorical optimization, our model’s decrease in AUC in Table 1 is expected. Figure 1 shows the learning curve of the KSV loss. Training with the existing classification loss \(L\) decreased KSV loss for some degree. However, when trained with the KSV loss, it reduces the loss dramatically. Figure 1 also shows the learning curve occasionally fluctuates. The weight added might be the cause: when the latest time step is weighted heavily, the target value changes more, making the loss fluctuate. Without adding them to the loss function and only training the model with our KSV loss, \(w_1\) and \(w_2\) dropped significantly on real datasets. For the Simulated-5 dataset, they increased possibly by overfitting to the accuracy rate. Since the Simulated-5 is a simulation dataset, it is different from real datasets. When the prediction overfits the accuracy rate, its changes between time steps become larger, causing waviness indicators to increase. We also examined AUC for the current time step in the column AUC \(C\). We noticed training the models with KSV loss has even less negative impacts on AUC \(C\).

In this paper, we addressed a new problem in DKT, namely the accuracy rate problem. To quantize and solve this problem, we proposed a new learning loss: KSV loss. The results show that our loss function is useful to measure how well the model fits the accuracy rate. Training with the KSV loss also reduced the waviness loss compared with the baseline. In the real world application, there are various ways of utilizing the model’s predictions. For example, a teacher may want to use the prediction results to assess students’ academic abilities, as is suggested in many research papers. With a more intuitive output of accuracy rate, our method is more suitable to be shown to students or teachers and opens up wider opportunities for KT models.

References

ITS Promoting Realization of Misguided Self-confidence in One’s Own Comprehension

Tomoki ABURATANI\textsuperscript{a}, Kazuhisa SETA\textsuperscript{a}, Yuki HAYASHI\textsuperscript{a} & Mitsuru IKEDA\textsuperscript{b}

\textsuperscript{a} Graduate School of Humanities and Sustainable System Sciences, Osaka Prefecture University, Japan

\textsuperscript{b} Graduate School of Advanced Science and Technology, Japan Advanced Institute of Science and Technology, Japan

*aburatani@ksm.kis.osakafu-u.ac.jp

Abstract: Most of the proposed intelligent tutoring systems (ITSs) intend for learners to cultivate a better understanding of what is written explicitly in texts regarding a target learning domain. However, these ITSs do not consider the improvement of learners’ awareness of their own understanding state. To solve this problem, we develop a method for building a learner model that captures learners’ misguided self-confidence in their own comprehension. More specifically, we set up a situation that allowed learners to proceed with presentation design activity using semantics-aware presentation materials (slides and terms of learning goals). In this situation, the materials have asymmetric structures between information that learners can recognize explicitly and computers can process. Based on these characteristics, our learner model allows ITS to capture what learners should set as learning goals, even if the contents are not described explicitly in the texts.

Keywords: Learner model, knowledge representation, presentation design tasks

1. Introduction

In higher education, the paradigm of learning should be transformed to orient with a structural understanding that comes from learning not only explicitly described knowledge but also tacit ones, such as the reason for the birth of a particular technology and its associated problems (Colthorpe et al., 2018). However, if teachers teach learning strategies without any concrete learning context, it is difficult for learners to ground such abstract concepts of strategies with their practical learning context such as Semantic Web (SW) (Bransford, 1999). In other words, a top-down instruction approach makes learners difficult to adapt these strategies to their learning contexts as an embodiment of learning goals that indicate their own evaluation criteria for learning achievement. Therefore, it is necessary to implement a bottom-up approach which grasps differences of criteria what they recognize they have already ‘understood’ particular concepts between a learner and others, and that encourages learners to evaluate their own criteria.

Many studies about intelligent tutoring systems (ITSs) propose strategies and systems that provide learners adaptive interventions for learning domains in which the knowledge, that should be understood, is defined clearly. The systems diagnose what knowledge the learners do not understand; therefore, there is a high degree of agreement about learning target domains (i.e., what should be understood) between the system and learners. Learning support methods which visualizes the learners’ own state of understanding, and which gives and make learners assemble components of the domain knowledge that should be understood, are typical types of this kind of system (Hirashima et al., 2015; Woolf, 2010).

In this study, we address the research question how to develop ITS that captures the learner’s state that learners think they have already understood even though others believe the learner does not understand yet, which we called ‘misguided self-confidence in their own comprehension,’ and how to encourage the learners to realize this state. To address the question, we propose (1) a task design which makes learners aware of learning goals that correspond with contents learners should understand, and (2) a system to intervene with the learners’ misguided self-confidence in their own comprehension.
2. Learning Task Design and System Architecture

2.1 Learning Task: Presentation Design Activity

In this study, we designed a learning task named Presentation Design Activities (PDA). For this task, we prepared slides (shown in Fig. 1(A-i-iii)) and a list of words that represent learning goals, namely ToLGs: Terms of Learning Goals (Fig. 1(A-i)) as presentation materials. ToLGs are concepts that express intentions of the presentation design, such as ‘make others understand the technical limitations of SW’ and ‘make others understand the differences between URI and URL.’ Learners are provided these presentation materials and required to select and sequence these prepared materials to represent what they think it is necessary to insert into the presentation. Hereby, they require to consider what they should make others understand to achieve the final goal of the presentation, such as answering the question, ‘what is SW?’ Learners then use the ToLGs, such as ‘make others understand that URI is a mechanism that gives unique identifiers to concepts and objects’ and ‘make others understand that URL is a mechanism that gives unique identifiers to places on the Internet,’ to embody in stages the achievement goals that they have set for each slide as pyramidal structure (Fig. 1(B)). This hierarchy should be structured to satisfy the following two points. First, the attached learning goals should be consistent with contents which are described in the relevant slide. Second, contents that is implicit should also be included in the learning goals.

The system captures these two states of a target learner and generates adaptive feedbacks. Based on this task, learners are encouraged to externalize and reflect on their own learning processes and notice the insufficiency of the rigor of their criteria of learning goals.

2.2 Learner Modeling and Feedback Generation

The system generates some feedback based on two modules. One is a module that captures whether the learners try to understand the content not only explicitly described contents but also implicit ones on each slide as a learner model (Learner modeler). Another is a module that generates some feedbacks based on the learner model (Feedback generator). The learner modeler processes learner models by comparing to knowledge sets that should be explained on each slide (Slide-K in Fig. 1) and knowledge sets corresponding with ToLGs which are assigned by learners to the slide on the pyramidal structure of presentation design (ToLG-K in Fig. 1). Then, the modeler calculates two lists as learner model. One is the OK list, which the learner thinks that should be explained from the intersection set between these knowledge sets. Another is the NG list, which the learner does not think from the difference set. Moreover, each list is divided into 2 lists based on consistent knowledge types explicit and implicit; therefore, the learner models is represented as a four type lists consisting of knowledge type ([explicit]-[implicit]) and the learner’s states ([OK (trying to explain)]-[NG (not trying to explain)]).

The system capture what concepts the learner remains unaware of the need to understand and what they should reflect on their criteria for learning achievement based on this learner model.

![Figure 1. Presentation Design Using Semantics-aware Materials.](image-url)
Table 1

<table>
<thead>
<tr>
<th>Content should be explained on the slide</th>
<th>Generated Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difficulties in balancing distributed semantic data development and consensus building on metadata definitions, and concrete examples</td>
<td>(1) To realize distributed and cooperative agents, it is necessary to specify computer-interpretable metadata, set up a common language among agents, and metadata of problem-solving procedures. You probably did not consciously try to understand this content explicitly written on the slides and (2) The realization of situational cooperative problem solving requires that different agents be able to exchange problem-solving situations with each other. (3) To deal with problems where problem-solving procedures are not programmed, a mechanism is needed to be able to read and use data from the problem-solving procedures. You probably did not consciously try to understand this content that is not written on the slides.</td>
</tr>
</tbody>
</table>

3. Pilot Study

We used the system in a learning context of Semantic Web (SW) to confirm the feasibility of what the system can be implemented in actual educational scenarios and generate feedbacks that contribute to realizing their misguided self-confidence in their comprehension. In this pilot study, we give learners a question, *What is SW, and what are the technical problems in its implementation?’* as an achievement goal of the learning. Furthermore, we prepared ten slides and 52 ToLGs. The test subjects are five undergraduate and three graduate students. They are majoring in informatics and had not been learned SW but they had acquired essential knowledge and skills which are necessary for understanding contents. We made them use the system and analyzed the generated feedbacks.

Table 1 shows an example of generated feedback. This example shows feedback to a learner on a slide, which describes identity problems in distributed data construction. More specifically, it is stated that “it is difficult to build consensus on some metadata among persons when some metadata are developed in a distributed manner as a SW technology” in the slide. Then, this feedback points out that the learner did not understand both written explicit content (1) and implicit ones (2), (3). This example shows that the system generated intended feedback for giving an opportunity to promote realization of the learner's misguided self-confidence in their own comprehension about (1) to (3) with appropriate natural language subjunctive clauses, and an appropriate diagnosis that captures what the learners should be aware of as learning goals. We have confirmed that all feedbacks demonstrated appropriate sentences same as this feedback, and the system worked without any problems in practical use.

4. Conclusion

In this study, we developed an intelligent tutoring system for promoting the realization of misguided self-confidence in learners’ own comprehension. Our system has a function that capture whether the learners try to set learning goals not only explicitly described but also implicitly expressed in texts, and the system also has a function that generating feedbacks for promoting realization of insufficiency of their criteria of learning goals based on presentation design activity. As future tasks, we plan to analyze whether the use of this system actually enhance to refine their recognition of self-confidence in their comprehension and to clarify the value from not only viewpoints of system architecture but also the viewpoints of educational theory.

References

Designing Technology Supported Scaffolding for Fractions Learning in Primary Schools

Daner SUN *a, Lan YANGb

a, b The Education University of Hong Kong, Hong Kong SAR, China
*dsun@eduhk.hk

Abstract: In the field of Mathematics education, there are the urgent needs of developing technology supported adaptions for students with learning difficulties due to the nature of Mathematical problem solving, the disparities in Mathematics achievement and the importance of Mathematics subject in education and job market. In this paper, a technology-supported learning environment with the employment of mechanism of scaffoldings: scaffolding changes and scaffolding logic for fractions word problems solving among students with LD was introduced. Specifically, the literature and rationales of the system design and development were discussed and presented.

Keywords: maths education, computer-mediated scaffolding, learning difficulties

1. Introduction

The term Mathematics Learning Difficulty is used broadly to describe a wide variety of deficits in mathematical skills. “Individuals who exhibit learning difficulties may not be intellectually impaired; rather, their learning problems may be the result of an inadequate design of instruction in curricular materials. Further research shows the use of technology is effective not only in helping students with learning disabilities in solving mathematics word problems but also beneficial to their learning in solving word problems with fractions and it can promote the inclusion of students with various disabilities (Fichten et al., 2009). The notion of scaffoldings has great potentials on relieving learners’ cognitive burden and with addressing the problem-solving process when integrated with appropriate pedagogical design, particularly, the computer-mediated scaffoldings (Quintana, et al., 2004). In addition, scaffoldings may relate not just cognitive skills but to other aspects such as emotive or affective factors (Yelland & Masters, 2007). Thus, we develop a technology-supported learning environment with the employment of mechanism of scaffoldings: scaffolding changes and scaffolding logic for fractions word problems solving among students with LD, which was few explored in Mathematics education. In this study, a web-based learning environment: ScaffoldiaMyMaths has been developed with unique features of adaptive scaffoldings. The rationales of the system design and development are discussed below.

2. Rationales of the system design and development

Scaffolding can be distinguished in four different dimensions, namely scaffolding scope, scaffolding intervention, scaffolding change, and scaffolding logic. In terms of scope, computer-mediated scaffolding can use either context-specific or generic scaffolding strategies due to theoretical models adopted by designers (McNeil & Krajcik, 2009). A generic scaffold is applicable to a wide range of units and context without changes in the scaffold itself (Stark, 2013). According to the theory of knowledge integration, context-specific scaffolding should be used and most previous studies in mathematical learning among primary school students have used context-specific ones (Hwang, Wu., & Chen, 2012) which are found to be effective. Based on these previous studies and the theory of knowledge integration, we will use context-specific scaffolding in this study. Second, there are four main categories of scaffolding intervention: conceptual, strategic, metacognitive, and motivation scaffolding (Hannafin, Land, & Oliver, 1999). Conceptual scaffolding highlights factors taking into consideration when solving a problem while strategic scaffolding proposes a target strategy like argumentation, problem solving or evaluation. Metacognitive scaffolding assists students to evaluate their understanding while motivation scaffolding enhances or maintains students’ motivation in learning like self-efficacy and autonomy (Belland, Kim, & Hannafin, 2013). Researchers have shown
the beneficial effect of conceptual and strategic scaffolding on learning mathematics among primary school students. Based on these prior studies, we will adopt both conceptual and strategic scaffolding in this study.

3. The Features of ScaffoldiaMyMaths Learning Environment

The web-based learning environment: ScaffoldiaMyMaths has been developed to facilitate students learning to solve fractions word problems. In ScaffoldiaMyMaths, the instructional lessons will be composed of five sets of word problems: where each set is consisted of 10 tasks with the same underlying structure (i.e. solution rationale) but different surface characteristics (i.e. cover stories and values). For each task, we design the scaffoldings accordingly. Explanation prompts in the form of work-out examples and problems (EP) as the key representative scaffoldings will be designed (Atkinson and Renkl, 2003). To help the student in identifying a problem solution, the solution steps will be visually and listed according to the nature of the task and the setup of scaffolding condition below. Below is the basic interface of the system.

![Figure 1. Student interface 1](image1.png)  
![Figure 2. Student interface 2](image2.png)

The design of the four adaptive scaffolding conditions:

1) **Fixed backward fading scaffolding condition**: In the first task, the solution of the work-out example will be presented to participants with all four solution steps sequentially shown to the participants. In other words, participants will be asked to work through the example step by step by clicking on the “next step” button and the next step will be shown to them on the top of the earlier step. In the final step, the four solution steps will be presented to them at the same time as a complete work-out example on the screen. After the participant complete the first task, they will continue to inspect the second task by clicking on a “next problem” button on the screen. The presentation format of task 2 is the same as the one in Task 1. The third task is similar to Tasks 1 and 2 except that the final solution step will be disappeared and participants will be asked to anticipate the final step by themselves and input the solution in the box to be provided, otherwise they are not able to continue. If the participant provides a wrong answer, the final solution step with the correct answer will be provided to them for their feedback. Similarly, the fourth task will be the same as the third one in presentation of the solution steps. In the fifth and sixth tasks, the last two steps will be omitted while in the seventh and eighth ones, only the first solution step will be displayed. Finally, in the ninth and tenth tasks, all solution steps will be skipped.

2) **Fixed fading and adding scaffolding condition**: In this condition, the scaffolding is fading from the first task to the fifth task, and then the scaffolding is adding from the seventh task to the tenth one. In other words, in first and tenth tasks, all solution steps will be shown, in the second and ninth tasks, only the first three solution steps will be displayed, in the third and eighth ones, only the first two will be presented, and in the fourth and seventh ones, only the first step will be displayed. Finally in the fifth and sixth ones, all solution steps will be omitted.

3) **Performance-based fading scaffolding**: In this condition, the basic schedule is the same as the one in the first condition: fixed backward fading scaffolding condition. However, if the participant provide a wrong answer to any task, the fading schedule will be delayed until the participant is able to finish two consecutive tasks correctly.

4) **Performance-based fading and adding scaffolding**: In this condition, the original schedule is also the same as the one in the first condition: fixed backward fading scaffolding condition. However, the scaffolding can be added or faded based on the performance of participants. In the first and second tasks,
if participants are not able to provide a correct answer to these tasks, the fading schedule will be postponed until they are able to complete two consecutive tasks correctly. The fading schedule will continue if participants are able to provide the correct answers all along. However, if they are not able to do that in any task after the fading has begun, the adding scaffolding will be initiated. In other words, the former omitted solution step will be presented again. If the participants continue to provide wrong answer, the adding will be continued until all solution steps are shown again. The online platform ScaffoldiaMyMaths (final version) is designed to be run in these four different conditions, which gives certain extent of scaffolding to learners in the experiment.

4. Conclusions and Further Work

The impact of the computer-mediated scaffolding approach could potentially be context-sensitive to the school settings and environments in Hong Kong. The usability test and pilot study of the system will be conducted stage by stage and the cognition research will be further conducted. In sum, the adaptive scaffoldings will facilitate teachers’ teaching students with less workload and more effective learning management. Most importantly, the well-tailored online lesson package with appropriate pedagogy will facilitate teachers’ teaching for students both in and out of classroom. It will offer students with more inquiry time and consolidate what they have learned in class by themselves rather than receiving or training the solutions directly by their teachers. The nature of technology learning environment could better provide the parents with students learning track and improve parent assistance of their children out of school, especially the free and frequent access of the system by the low-income family.

Reference

Unpacking students’ modelling behaviour in the Sun-Earth system: Use of digital media tool-based epistemological resources

Hinal KIRI*, Harshit AGRAWAL*, Shanize FORTEb, DurgaPrasad KARNAMa & Sanjay CHANDRASEKHARANa

*aHomi Bhabha Centre for Science Education, Tata Institute of Fundamental Research, Mumbai, India
bD.G. Ruparel College of Arts Science and Commerce, Mumbai, India
*hinalkiri@yahoo.com

Abstract: Understanding scientific models described using external representations is an important objective of science education. However, modelling and reasoning using models are not easy for students due to inaccessible phenomena or abstracted/idealized nature of the models. Learners rely on various epistemological resources to make sense and meaning of the models in any given context. However, the underlying processes and the interplay of these resources in the process of constructing and reasoning with these models are not completely understood. We use the case of the Sun-Earth system to unravel these dynamics, by a detailed study of 4 undergraduate (UG) science students, using questionnaires and interviews around interactions with a digital media tool (designed by us). The results throw some light on the dynamics of students’ use of various epistemological resources (lived experiences, book-based, and tool-based experiences) in their process to construct models and reason with them. We also illustrate, in this process, a potential role of digital media interfaces in providing epistemic access to the inaccessible by being an epistemological resource of a new kind.

Keywords: modelling, digital media, sun-earth system, epistemological resources

1. Introduction

Models are used in science to describe or explain a physical phenomenon, by standing for relation with the phenomenon. They capture the structural relationship (five structure-types (Hestenes, 2010)) between the various components of the physical system. These components of a physical system (e.g. idealised moving mass) are abstracted or idealised features of the physical phenomenon (e.g. mass or acceleration as abstracted components of a real moving object, and the object idealised as a point mass) (Suppe, 1989; Suppes, 1960). These models take the form of external representations: physical models, spatial representations like diagrams, linguistic or mathematical formulations like statements or equations, simulations etc.

From a science education perspective, it is often difficult for learners to make sense of the epistemic role of the models and their relationship with the modelled. Furthermore, the process of systematic modelling (creating, testing and revising models as a community), a defining practice in science (Giere, 1988; Nersessian, 2002), does not reflect in the science learning (Lehrer & Schauble, 2010) in our current educational systems. Besides limited understanding of the nature of science among teachers, epistemic inaccessibility could be a key source of difficulties for the students in understanding and using the models effectively. The phenomena that are modelled in the science are often not immediately accessible to our sensorimotor system (e.g. phenomena at the subatomic level or cosmological level) (Pande, 2018) and the models themselves often represent abstracted or idealised components (minimised by removing non-relevant features e.g. texture or colour of the mass); hence they are not epistemically accessible.

Shifts in educational research from a deficit (like misconceptions) to a resources model (epistemological resources) (e.g. Hammer & Elby, 2003; Smith et al., 1994) acknowledges students' lived experiences of these phenomena, among others, as a valuable resource in their reasoning and
learning. The experience of students with models of these phenomena through textbooks and classroom discourse can also be seen as a kind of epistemological resource. It is reasonable to assume that learners would deploy various epistemological resources at their disposal in the process of learning about or understanding a scientific model. However, there is very limited understanding of the underlying processes and the interplay of these resources in the process of constructing and reasoning with models. This study attempts to unpack these processes using a particular case of the Sun-Earth system (elaborated in section-2).

Towards this objective of unpacking the modelling behaviour and the interplay of various epistemological resources in the case of the Sun-Earth system, we use a digital media tool as a probe and as a tool-based epistemological resource. We take advantage of the potential of digital (or computational) media (Balacheff & Kaput, 1996; Papert, 1980) in enabling a better understanding of models by affording (Gibsonian sense (Heft, 1989)) novel sensorimotor interactions, thereby providing epistemic access (Karnam et al., 2019) to the models (epistemic availability (O’Donovan-Anderson, 1997)). We describe the digital media carefully designed to enrich experiences and trigger modelling behaviour in learners’ about the Sun-Earth system in section-3. We report a study (in section-4) involving a detailed examination of the modelling behaviour (and use of various epistemological resources) in learners (UG Physics students) triggered by the digital media design. This examination would throw some light on the nuances of the underlying dynamics, possibly provide novel perspectives for learning scientists that have implications for future digital media designs.

2. Learning of Sun-Earth system

The Sun-Earth system provides an apt context to study modelling behaviour for at least 3 strong reasons. Firstly, astronomical phenomena have always fascinated the human mind and historically appears to have been among the first that humankind modelled in nature, as reflected in cultural narratives about astronomy across civilizations. For us, the Sun-Earth system provides a very rich context to unravel the modelling behaviour, because of its direct visuospatial and dynamic nature; the model of the Sun-Earth system is a model of the spatial and dynamic (moving) relation of the elements (the Sun, the Earth) in space. This direct one-to-one mapping between the elements of the model and the phenomenon significantly simplifies the complexity involved in unpacking the modelling processes in human minds.

Secondly, the elements of the Sun-Earth system have a strong real-world context to the learners and hence an attribute of concreteness. Learners across age groups, in their daily physical interactions, encounter and have meaningful relations with these elements. Further, the related phenomena of day-night and the seasons have direct and immediate salience and hence are perceptible to us. This is unlike other cases like atoms or plant/animal cells, which are epistemically distanced from the learner’s lived experiences. We assume that it is more probable for a learner to have meaningful real-life encounters with at least the Sun (if not with the Earth) as an entity, than with, say plant cells or electrons in an atom of an object; even if they interact with objects and plants, the notion of an atom and cell as an entity, is still inaccessible. Thus, this topic is accessible and simple to study due to the lived experiences of the modelled phenomenon.

Lastly, from a more practical standpoint, the Sun-Earth and moon system is a topic well studied in education research (perhaps this was in turn because of the above two reasons). The studies range from those which highlight the alternate conceptions to those analysing the conceptual aspects of the system and those providing various interventions. The section-2.1 provides a quick review of the existing literature on the aspects explored about the Sun-Earth system and related phenomenon (like day-night and seasons) and the existing ways of teaching and learning this topic and the interventions. The moon could also have been incorporated in our study, but we chose to stick to the simplest model and simplest phenomenon to look at the process’s underlying modelling behaviour and interaction between various aspects (formal representations, real-world experiences etc.) and creation of models.

2.1 A brief review of research related to the Sun-Earth like systems
Students are found to hold numerous models about the Earth (five mental models like being in the Earth, flat Earth, out of Earth etc.), Sun-Earth system (versions of geo-centricity) and related phenomena like day-night (Chiras, 2008; Samarapungavan et al., 1996; Vosniadou & Brewer, 1992, 1994) and seasons. The Sun-Earth-Moon system (Jones et al., 1987), Earth and space (Schoon, 1992) and overall astronomical understanding of children (Baxter, 1989; Schoultz et al., 2001; Trumper, 2001) have also been widely examined. Numerous interventions have been suggested in various forms such as (1) learner cognition driven ones like those based on inquiry-based learning (Salierno et al., 2005), mental model-based strategies (Taylor et al., 2003), gestural instruction (Padalkar & Ramadas, 2011); (2) technology-based ones like AR and VR interventions (Bakas & Mikropoulos, 2003; Shelton & Hedley, 2002); (3) pedagogical/ course interventions as diverse as conversation-based ones about the moon with Japanese teachers (Suzuki, 2003), engineering design-based engagement (Dankenbring & Capobianco, 2016) (given the recent push for engineering sciences in the US), among others. However, a careful and detailed explanation of students’ current understanding from a modelling perspective is not satisfactorily provided. This study of unpacking modelling behaviour in the context of the Sun-Earth system could thus be a mutually useful and insightful exercise to the audience of astronomy education and audience interested in modelling in general.

3. The Digital Media tool

The standard Sun-Earth model geometrically captures the relative position and motion of a tilted Earth and the Sun. This models the movement of the Sun as observed from Earth and other phenomena like day-length, shadow and seasons. It is often difficult for learners to imagine the spatial dynamics and meaningfully linking the model with the observed phenomena using the static pictures in books. Digital media tools (hereafter referred to as just tools, unless specified otherwise) can address both the difficulties by dynamically linking multiple perspectives. They also afford experimentation and exploration of different scenarios, thus enriching their modelling experiences of phenomena, which are often epistemically inaccessible.

Different tools focus on different parts of the difficulty with understanding and reasoning about the Sun-Earth system. For example, ‘Sun Position’ app on Google Playstore shows the Sun’s trajectory in Augmented Reality; ‘3D Sun-Path’ shows the 3D trajectory of the Sun as viewed from different places on Earth (http://andrewmarsh.com/software/sunpath3d-web/). Both these systems allow exploring the phenomena but do not directly support developing an explanatory model. The Season’s Simulator (NAAP) provides multiple perspectives of the Sun-Earth model; it supports basic causal reasoning but is not sufficient to build and explore the geometrical model. To suit our modelling requirements, we custom-designed a tool informed by design recommendations based on multiple external representations (e.g. Pande, 2018; Virk et al., 2015) trying to address the above shortcomings.

3.1 The Design

The design requirements were to provide affordances to explore phenomena in multiple perspectives, which are tightly interlinked. The design is broadly informed by the 4E models of cognition, which emphasize the ‘constitutive’ role of sensorimotor interactions between the body and the environment in shaping the cognition (and imagine using models in our case) (e.g. Glenberg, 2010; Hutchins, 1995; Sterelny, 2004; Thelen & Smith, 1996; Van Gelder, 1999). The design should be able to provide meaningful experiences of the usually inaccessible phenomena and the underlying abstract models. These experiences along with other experiences can thus trigger the imagination of the dynamics, and facilitate grounded conversations about the intricacies of the Sun-Earth system and the related phenomena.

The tool was built using the Unity Engine. The NREL’s Solar Position Algorithm (SPA) was used to provide the values of azimuthal and incidence angles for a given position and time. It has 3 different interlinked views (see figure-1) presented in 3 view-panels.

- The first view or orbital view (top): This panel dynamically captures the geometric constraints of the model showcasing a tilted Earth rotating around its axis and simultaneously revolving around the Sun. The Sun is static and the Earth’s axis and path are displayed. This perspective typically
The second view or out-of-Earth view (bottom left): This panel takes the position of an observer hovering just above the Earth, which was not present in the tools we came across. The Earth’s axis, the normal at a given place, the sunrays and the geographical north are marked as arrows. This acts as a bridge between the faraway orbital (first) view, and the on-Earth (third) view, connecting which could be difficult due to the differences in the scale.

The third view or on-Earth view (bottom right): This panel corresponds to the lived perceptual experience of seeing the Sun move across the sky. The geographical directions, the sunrays and the normal are displayed as arrows, and the Sun’s trajectory is shown by a yellow arc across the sky. This view with a house and changing shadow could connect to the learners’ lived experiences.

The interface has controls for changing the various parameters involved. Users can change the location on Earth for which the observations have to be made. The date and time can be changed to account for changes across the year. A play button was added to automatically increment the hour (Figure-1).

In the first view (top) the orbital position of Earth changes with a change in the day of the year. The axial tilt remains constant throughout; making it easy to observe that the northern and southern hemispheres are not equally lit throughout the year. This would be crucial for making sense of the differences in the Sun’s trajectories and seasonal patterns in the northern and southern hemispheres. This view emphasizes that if the Sun is taken as the reference, the sunlight incident on Earth does not change its direction significantly in a day (24 hours). Therefore all the changes in the Sun’s position during the day must be due to a change in Earth's rotation. Appropriate lighting effects ensure that half of the Earth facing the Sun is lit up while the other half is dark. The third view (bottom right) shows the position of the Sun in the sky at any given location, time, day of the year. The trajectory of the Sun for the entire day is also visible. This makes the change in the Sun's trajectory easier to observe. According to the change in position of the Sun, the shadow cast by the house also changes. This view supports panning, rotating and zooming to provide rich nuanced observations. The second view (bottom left) shows the corresponding location on Earth.

Transitioning between views happens via carefully designed inter-perspective elements. The part of the Earth lit in the top view is the same that is lit in the bottom left (second) view. The purple line shows the direction of the sunlight in second and third views. Two green arrows in the same view show the normal at that location and the north direction. These 3 lines are also provided in the bottom-right (third) view. Further, the Sun’s position is seen from the on-Earth (i.e. third) view which also supports panning, zooming and rotating. These views and the inter-perspective elements change in real-time with the changes in the time, date and location, providing dynamically linked multiple perspectives.
Further, the three views together provide powerful investigation (exploration and experimentation) opportunities for the learners. For example, figure 2 shows a scenario. On June 21 (summer solstice), at noon in Mumbai, the second view indicates the angle of sun rays (purple) with the normal (green). This angle is a direct result of the position of Earth in the orbit around the Sun and its rotation around itself. The time and date parameters in the first view determine the angle of the Sun’s ray in the second view. As the time is changed, the Earth can be seen rotating in the first view, the direction of sun rays (purple arrow) can be seen changing in the second view. The third view extends the second view by providing the on-Earth perception of the above changes; the changing angles in the second view are directly reflected in the third view. Thus, the position of the Earth in its orbit and its angle of the rotation completely determine the Sun’s position in the third view. This was one of the design requirements we started with.

These features of dynamically interlinked multiple perspectives affording powerful investigations can provide well-structured and rich modelling experiences to the learners. This thus can act as a new kind of epistemological resource, which we refer to as a tool-based resource, in this paper.

4. The study

The study aims to unpack the interplay of various epistemological resources in students’ modelling behaviour. We used the above tool along with a set of questionnaires and interviews to trigger and examine modelling behaviour in students about the Sun-Earth system.

4.1 Methodology: Participants, Material and protocol

The participants were 4 UG Physics students (S1-S4). They were administered a written test (with an opportunity to describe their responses using text as well as diagrams). This was followed by a series of tasks with our tool and some conversations during the tasks. The participants worked in groups for these tasks. Each group spent about 1.5 hours interacting with the system. After 2 weeks, they were administered the same written test again. They were interviewed (students S1, S2, S3 together and, S4 one-to-one due to logistical constraints) in the context of their test responses to capture their modelling behaviour and the interplay of various epistemological resources. The usage of the pre and post-tests allowed us to capture changes in their modelling behaviour longitudinally and gave a better grasp on the dynamics in their reasoning supported by the tool, as an additional epistemological resource.

The tests had questions (they can explain using text and diagrams and later elaborate during the interviews) exploring their reasoning about the day-night lengths, seasons and the changes in them with time (of the year) and location on Earth. The tasks on the digital media tool were exploratory and open-ended and tried to put the students into situations that triggered their mental models about the motion of the Sun. Some of these tasks were: (a) observe if the direction of sunrise and sunset changes for a given location through the year; (b) is the Sun directly overhead at noon on all days? Create a directly overhead Sun for some locations; (c) look for patterns of change in the trajectory as we move from the equator to the poles.
4.2 Analysis Framework

The written scripts and the conversations recorded during interviews were analysed for the modelling behaviour and interplay of various epistemological resources. This involved tracking the usage of various epistemological resources in their reasoning through conversations. To do this, we sliced the data into streams of conversations called lines of reasoning (LOR). Any continuous flow of utterances anchored to a particular context or topic of conversation can become the unit of analysis, the LOR. These lines of reasoning (LORs) are dynamically constructed in real-time by an individual in the context of the conversation deploying various epistemological resources; these correspond to underlying processes of construction and manipulation of mental models (aspects of modelling behaviour). These dynamics are inferred from the sets of articulations (drawings and words) and gesticulations (gestures) as they reason and communicate. The epistemological resources for learners in our case could be broadly categorised as:

- **Lived experiential resources**: from students’ lived experiences with Sun-Earth - indicated by references to their experiences (e.g. observations of Sun, shadows while travelling or at their house or colleges etc.) and related extensions.
- **Book-based resources**: from formal sources of science - indicated by references to textbooks, etc. (e.g. use of terminology or diagrams in their textbooks or scientific discourse etc.) and related extensions.
- **Tool-based resources**: the experiences while doing tasks on the tool - indicated by references to the elements in the tool and the experiences with the tool and related extensions.

Different resources can play different kinds of roles in shaping their reasoning in a particular situation and reflect in their modelling behaviour. Here, we could operationalise desirable modelling behaviour as an ability to coherently use and apply different resources, and progressively create, test and revise one’s mental models. This locally coherent integrated model is referred to as an epistemological frame (Elby & Hammer, 2010).

The quality of the students’ modelling behaviour is tightly linked to the coherence of underlying mental models that are dynamically constructed, tested and manipulated as they reason. This is reflected in the coherence of LORs — the meaningfully connected streams of utterances in the conversations. Students put epistemic efforts to ensure this coherence, deploying multiple resources at their disposal, striving to resolve any inconsistency or cognitive conflicts. So, coherence of LORs can be a good indicator of the quality of modelling behaviour and any personal reconciliation can be a potential marker of meaningful learning.

5. Analysis, Findings and Discussion

We analysed the post-test interviews and the written scripts of the students for patterns in their modelling behaviour by transcribing the videos with descriptions of utterances (speech, gestures and drawings). These are iteratively organised into meaningful streams (series of episodes) of conversations as LORs. Then we reflected and inferred possible epistemological resources applied in this reasoning process based on the indicators previously outlined. There are certain situations, where the learners could engage in a conversation with coherent lines of reasoning. And there were certain other situations where they became less systematic and appear to lose track of a LOR; such situations had cases where they were throwing scientific terms (from book-based resources). We report some of such episodes, in detail, unpacking the dynamics below. In this paper, we confine to illustrating an initial application of the above analysis framework to some episodes, indicate some patterns of the interplay of various epistemological resources and highlight the usage of digital media interface as a tool-based epistemological resource; detailed analysis is in progress.

5.1 Deployment of various epistemological resources in Coherent Modelling Behaviour

While explaining the formation of zero-shadow (or overhead Sun position), S1 indicated the Sun rising and setting with his large semi-circular hand movements from one end to the other (indicating his current imagination) (see Figure 3). A corresponding explanation was also found when he drew the
typical semi-circular figure from their textbook (see right of figure-4). S1’s gestures can be ascribed to interconnecting the book-based resources (diagram) as well as the lived experiential resources of typical Sun’s trajectory.

Both S1 and S4 used the angle between the normal and the sun rays (tool-based resources in the third view) in their reasoning, as indicated by gestures (palms orientated in a particular way corresponding to the arrows) as well as the drawings (Figure 4). These two arrows are found to provide them with very strong conceptual tools to reason about the formation of shadow (something to anchor upon similar to the attentional anchors (Abrahamson & Sánchez-García, 2016)). Here they could apply various epistemological resources (including tool-based ones) effectively. The LOR was coherent as long as they reasoned about the shadow in a given location.

**Figure 3.** S1 showing Sun’s movements- Sun at the top (left); Sun moving from East to West (his left to right) (mid); Sun setting (right).

**Figure 4.** S4 showing using his palms (left) and S1’s response in the test (right) indicating using normal and the direction of Sun rays in their reasoning.

### 5.2 Deployment of various epistemological resources in Incoherent Modelling Behaviour

However, an extension of the above LOR — about zero-shadow in a given location — to different locations did not happen effectively. For example, S3 said zero-shadow will happen everywhere at least on one day in the year. When we pursued this LOR, it started getting incoherent. He then employed the conceptual tools of the normal and sun rays (tool-based resources, which has worked in the last situation) in an attempt to get a grip on it. S2 intervened and said ‘this does not happen everywhere, but only near the Tropic of Cancer’ (recollecting the experiences during the tasks hence as a tool-based resource as well as a possible book-based resource due to the reference to the Tropic of Cancer). This LOR is shaky and incoherent as reflected in S2’s tentative statements like ‘(near the tropic of cancer) the angle (between normal and Sun’s rays) is less and (hence) higher chances (for zero shadow)’. Here we can see the tool-based resources deployed to some effect, but the LOR is not very coherent yet, and nor is the underlying mental model.

Later, when they tried to resolve the confusion, S3 brought in the idea of equinoxes (book-based resources). He tried to make a connection between the longest day and the Sun being exactly overhead (or zero-shadow) and when asked, he was able to justify by gesturing and saying that the Sun rises exactly in the East and sets exactly in the West (his gestures indicated he was meaning the plane of the Sun’s trajectory does not tilt towards North or South directions). Here, he was deploying the tool-based resource – the visualisation of the Sun’s trajectory in the third view – in his LOR. But, he too could not coherently extend the reasoning to the shift with latitudes and used terms like ‘zenith’ incoherently. In a similar situation when LORs broke, S4 (interviewed one-to-one) too used terms like equinoxes and solstices (both of them are book-based resources).

Eventually, jumping some steps of reasoning, S2 and S3 concluded that the zero-shadow is possible only on the Tropic of Cancer and Tropic of Capricorn. A break in the LOR was evident when S1, who was there in the discussion but could not participate, said – “actually when there was the tool (earlier), I understood clearly. But now I am confused, as a lot of terms are used”. After some reflection,
he attempted to reconcile the break in the LOR by recollecting the task using the tool: with Mumbai as the location, by changing the days, the plane of the Sun’s trajectory was coming closer and farther from the normal, and they could not get an exact zero shadow. Here he could sense a broken LOR and when explicitly asked, tried to reconcile using tool-based resources. Though the tool helped in reconciling the broken LOR, the mental model is still incoherent; this indicates the way a digital media design could provide epistemic access by providing a new kind of epistemological resource and triggering imagination, especially useful for the learners.

To explain seasons, they continued the LOR from overhead shadow (with the Sun’s trajectories) and connected it with the day-length (long days in summer and short days in winter—looking at the changing length of the Sun’s trajectories: tool-based resources). S3 extended this by bringing in resources from the topic of heat and explained, the days are longer and hence more heat. However, when we asked S3 on how he connected this with the elliptical orbit (another representation used to model, that he drew in his written script), he drew the conventional Sun-Earth elliptical orbit diagram (book-based resource) and gave the distance-based explanation (Earth being close to the Sun in summer and farther away in winter). At this point, all of them were confused in explaining different seasons in the hemisphere at any given time and fell back on the Sun’s trajectory-based description for explaining. This captures interesting friction in the process of testing and integration of resources from two different sources about the same physical phenomenon. They could not coherently construct the LOR starting from the changes in Sun’s trajectory from the third view of the tool to the seasons interconnecting the tilt of Earth, and eventually fall-back on the distance-based explanation.

![Figure 5. S1 and S2 explaining 6-month long day-night at poles.](image)

### 5.3 Other observations

See Figure 5. In another conversation related to day and night at poles, S1 explained using a pen in a fist as Earth and its tilted axis and shows the way the North Pole receives sunlight for 6 months and then the South Pole (which was discussed while using the tool). S4 indicated the Sun’s trajectory with his finger precisely explaining the 6 month-long daylight at the pole (North). This is something that they have tried on the tool (tool-based resources). Interestingly, the lived-experiential resources were not indicated much. So, when explicitly asked how much they used their lived experiential resources when answering these or during the 2-week break after their interaction with the tool, S1 and S4 said that they had observed the phenomena of shadow (overhead Sun) and sunrise time earlier too but never bothered to pursue them further and connect with what they were taught in their schools.

### 6. Conclusion

The above episodes of both coherent and incoherent LORs attempt to illustrate unpacking of the learners’ modelling behaviour revealing the complex interplay of various epistemological resources to build coherent mental models to reason with. In the above processes, the episodes of confusion like the extension of the ‘overhead shadow’ LOR to different locations on the Earth or the friction integrating book-based resources and tool-based resources to explain seasons, needed more intervention beyond the limited conversations during the interviews. If interventions enhanced by the tool embedded in a continued discourse are ensured, the learner could test, revise their model or create an entirely new model for themselves, which can be considered as successful reconciliation and meaningful learning.

The paper also illustrates how digital media interfaces can be effectively used to provide epistemic access to — by providing conceptual tools (like the 2 arrows: normal and sun rays in 2nd and 3rd views of Figure-1) — and triggering imagination of otherwise inaccessible models. From the diverse
deployment of tool-based resources in the above episodes, we illustrate the affordance of digital media interfaces providing new kinds of experiences and epistemic access to the inaccessible phenomena and models, and being a new kind of epistemological resource to the learners. Further, in a literal sense, the tool aligns with ‘perspective-taking’ involving the notions of ‘diving-in’ and ‘stepping-out’ (Ackermann 1996).

As we highlight the potential of digital media, we would also want to clarify the need for wider contexts integrating various epistemological resources. These contexts could emerge through more discourse in resolving the learners’ confusions. This has practical implications to educational technology designers, to consider technological interventions as a part of a holistic intervention integrating other epistemological resources that learners already come with, and not be thought of as a one-pill solution for learning problems, akin to a technocentric critique (Papert, 1987). The study acknowledges that by being part of a community of peers (with diverse epistemological resources), a rich and meaningful discourse could emerge fostering the desired modelling behaviour with richer models, and digital media systems can be one among various epistemological resources. Furthermore, our study, in a way, helps deepen the existing descriptions of such reconciliations using various resources including cultural knowledge systems (e.g. Jegede & Aikenhead, 1999).

This paper, thus, in its limited scope, tries to illustrate unpacking of the modelling behaviour, using episodes of reasoning about the Sun-Earth system. This describes the underlying interplay between different epistemological resources in shaping the LORs based on models. For researchers in the learning sciences, this unpacking of the episodes could be insightful. We hope this paper could contribute to systematic conversations around learning dynamics and to the role of technological interventions in meaningfully supporting learning.

The tool as well as the study are preliminary and have certain limitations. Some reflection of this can be seen in student’s post-questionnaire feedback to visually denote and mention numeric angular values (azimuth and incidence) in the out-of-Earth and on-Earth views. Also, the suitability of the design in connection to the existing formal school education is not accounted for in this paper. The study though gets sufficiently deep to deploy the analysis framework, a more rigorous data collection and application of the framework in future could give promising insights about the modelling behaviour, and the role of digital media tools in enriching learners’ experiences. Also, more explorations of learners’ modelling behaviour in spatial contexts with multiple perspectives other than the Sun-Earth system can strengthen the generalisability of the insights.

References


Lehrer, R., & Schauble, L. (2010). What kind of explanation is a model? In Instructional explanations in the disciplines (pp. 9–22). Springer.


From Monitoring to Sharing of Attention in Dyadic Interaction: The Affordances of Gaze Data to Better Understand Social Aspects of Remote Collaborative Problem Solving

Johanna PÖYŠÄ-TARHONEN\(^a\), Nafisa AWWAL\(^b\), Päivi HäKKINEN\(^a\) & Suzanne OTIENO\(^c\)

\(^a\)Finnish Institute for Educational Research, FIER, University of Jyväskylä, Finland
\(^b\)Assessment Research Centre, ARC, University of Melbourne, Australia
\(^c\)Department of Psychology, University of Jyväskylä, Finland

*johanna.poysa-tarhonen@jyu.fi

Abstract: This paper aims to better understand the social aspects of collaborative problem solving (CPS) through studying joint attention behaviour (JAB) in an online game-like environment. To capture these behaviours and exemplify how ‘jointness’ is achieved in CPS in remote dyadic interaction, event-related measures are utilised based on the following multiple interaction data: (1) individuals’ gaze data from CPS task completion and (2) automatically generated log files (i.e. chats and actions) from dyadic interactions. The results give empirical evidence of the detached, individualistic attention experiences (i.e. monitoring and common attention) and of bidirectional relations (i.e. mutual and shared attention) in which partners adopt an engaged approach towards one another to solve the task together. It is also observed how lower level attention in CPS can be a precursor to a higher level; that is, during interaction, there is a move from monitoring the partner’s actions towards common attention experience. In addition, it is noticed that richer second-person relations may come in degrees. In methodological terms, the gaze data can provide access to better uncover dyadic processes during remote CPS, but without the information embedded in the log data, they would not provide sufficient contextual details of the real interaction to fully understand social connotations related to CPS.

Keywords: collaborative problem solving, joint attention behaviour, live eye tracking, process-orientated research

1. Introduction

Collaborative problem solving (CPS), a key competence of 21st-century learners, is defined as a skill set required for solving problems in non-routine, collaborative situations in different domains (e.g. Funke, Fischer, & Holt, 2018; Graesser et al., 2018; Scoular & Care, 2020; Scoular, Care, & Hesse, 2017). In most frameworks, CPS is based on a socio-cognitive approach to learning, and it is seen to lie in a two-dimensional space of social and cognitive components that intermingle over the processes of problem solving (e.g. Funke et al., 2018; Graesser et al., 2018; Scoular et al., 2017; Zwiecki, Ruis, Farrell, & Williamson Shaffer, 2020). However, as the theorised CPS constructs are relatively new and rooted in individual problem-solving approaches, the social components have not yet been fully covered in the existing CPS models (Funke et al., 2018; Scoular et al., 2017). Therefore, endeavouring to understand how social components function in CPS is an essential step in moving from individual problem solving in a social context to a CPS construct in which the social and cognitive components are more amalgamated.

To better understand CPS and the qualities related to its social components in dyadic interaction, this study applies the concept of joint attention behaviour (JAB; e.g. Carpenter & Liebal, 2011; Korkia Kangas, 2018; O’Madagain & Tomasello, 2019; Seemann, 2012; Siposova & Carpenter, 2019; Tomasello, 1995), the foundation of any interaction predicting productive collaboration (Barron, 2003). However, despite the growing interest in studying joint attention and its premises to understand dyadic
interaction, no unified interpretation exists for what is considered joint attention and how ‘jointness’ in joint attention is achieved (Carpenter & Liebal, 2012; Seemann, 2012; Siposova & Carpenter, 2019). As a recent viewpoint to better understand the complexity related to joint attention, Siposova and Carpenter (2019) propose an approach to joint attention and social knowledge as a process of closely connected yet distinct phenomena of social attention. They argue that, instead of a single ‘state’ of joint attention as a binary event (i.e. there is jointness or there is not), jointness may come in degrees.

To capture behaviours and exemplify how ‘jointness’ and common knowledge in JAB is achieved in CPS in remote settings, event-related measures based on multiple interaction data (gaze data, logfiles) are utilised. To do this, the study takes the theorised CPS construct (Hesse, Care, Buder, Sassenberg, & Griffin, 2015) and the unique properties of JAB in dyadic interaction in a remote, game-like CPS assessment environment (Assessment and Teaching of 21st Century Skills [ATC21s], http://www.atc21s.org; e.g. Care, Griffin, & Wilson, 2018; Care, Scoular, & Griffin, 2016; Scoular et al., 2017) as its point of departure. Here, a student collaborates with another student, and the collaborative tasks aim to stimulate and elicit the social and cognitive elements and sub-elements that are part of the complex CPS framework by Hesse et al. (2015). In this environment, the level of participation can be compared with a real social situation with dynamic stimuli (i.e. actionable artefacts) and a chat property as the communication affordance for dyadic interaction. While the automatically generated logfiles as chat and actions of interacting dyads incorporate multiple information of joint processes, to make visible the typology of jointness as defined by Siposova and Carpenter (2019), the gaze patterns—recorded with remote eye trackers from the individual partners—are seen as significant. To better understand JAB in CPS, we expect that focussing on these gaze patterns in parallel with the interaction sequences of the communicating partners identified from the activity log data will help us ‘go beyond’ these sequences (Korkiakangas, 2018) and better understand CPS in this regard.

**A Typology of ‘Jointness’ in Joint Attention Behaviour**

Generally, joint attention is defined as a capacity to focus together with another on an external source or object in the environment (e.g. Eilian, 2005; Korkiakangas, 2018; O’Madagain & Tomasello, 2019). According to Siposova and Carpenter (2019), the objects of joint attention can be diverse sensory inputs, such as visual or auditory stimuli, or they can be present, past or future events or even mental states (i.e. ideas, plans). Given that gaze following is viewed as a promising basis of JAB (Seemann, 2012), it is also seen to include the coordination aspect of joint attention and the sharing of attention (Carpenter & Liebal, 2012; Tomasello, 1995). In its richest definition, individuals must equally recognise that they are attending to the same thing (O’Madagain & Tomasello, 2019; Siposova & Carpenter, 2019; Tomasello, 1995). Thus, following Carpenter and Liebal (2012), it is only communication that ‘turns mutually experienced event into interaction, into something joint’ (p. 168).

To better understand the complexity and multiple definitions of JAB and how it is achieved, Siposova and Carpenter (2019) argue that joint attention should not be considered a single state but as a process comprising various closely connected but distinct phenomena that can be discovered in related literature. Accordingly, Siposova and Carpenter (2019) have developed a spectrum of ‘jointness’ described as ‘a typology of social attention and social knowledge’ that aims to cover the diversity of the existing definitions. The typology comprises four ‘levels’ (basic components of JAB)—monitoring, common, mutual and shared—that all include the notion of a triadic relationship between self, other and an object or their attention. As a precondition for each of the four levels of social attention is the individual’s ability to engage in individual attention.

The first level in the spectrum of jointness is called **monitoring attention** (Siposova & Carpenter, 2019). This refers to a situation in which an individual is taking an observer’s perspective on a second individual involved, and in this way, attending to the same matter to which the partner is attending. At this level, the participants have individual knowledge of the situation, and their attention levels are independent; at the same time, an individual has knowledge that the other participant is paying attention to the same object or situation. Nevertheless, although both individual participants are simultaneously monitoring each other’s attention to the object or situation, they still assess the attention and knowledge states of the other participants individually. Often, monitoring behaviour is observable, such as turning one’s gaze or bodily orientation, but such behaviour can also be present without easily noticeable actions.
At the second level, *common attention*, two individual participants take an observer’s perspective, and nearly simultaneously, attend to what the other is focussed on (Siposova & Carpenter, 2019, p. 262). Here, individuals not only attend to the same object or situation but are also *attending to each other’s attention* to the object or situation. Engaging in common attention requires that the object of attention is pronounced and marked; that is, the participants can both assume that they are attending to the same object or situation. In addition, they have a reason to consider other participants’ attention; for example, they have a *predefined common goal to be achieved*, and in this respect, for both participants, the other individual’s attention is relevant. Siposova and Carpenter (2019) point out that, ‘under these conditions individuals could know they are attending to each other’s attention without any contact or communication’ (p. 262). The dependency of the other at this attention level is based on the awareness that they are both engaging in the same attention processes. Yet, notably, the *evaluation of whether they are in common attention is based on an individual’s perspective*, and thus, it may not be correct.

According to Siposova and Carpenter (2019), at the third and fourth levels of social attention, *mutual* and *shared attention*, the observer’s attitude toward the other and his/her attention no longer exists (i.e. a third-person experience), but the experience is based on direct commitment to the other, where the participants are both senders and receivers of the information (i.e. a second-person experience; see also Zahavi, 2015). Through direct social interaction, each participant becomes a ‘constituent part’ of the experience of the other (Zahavi, 2015), and attention to an object or situation is coloured by mutual awareness of each other’s attention (Siposova & Carpenter, 2019). This bidirectional nature makes the experience different if compared with monitoring and common attention levels that are individualistic (Siposova & Carpenter, 2019). Thus, in *mutual attention*, the participants are more or less simultaneously attending to the same object or situation but not necessarily communicating intentionally (Siposova & Carpenter, 2019). If compared with common attention, at this level, their experience is co-created.

The fourth level of social attention, *shared attention*, meets the qualifications of mutual attention, but this level also requires the participants to *deliberately communicate* with each other about the object or situation and/or the fact that they are sharing attention to it (Siposova & Carpenter, 2019). Thus, what makes shared attention different if compared with mutual attention is its intentional nature. Shared attention is characterised by behaviours in which individual participants verify to each other they are attending to the same object or situation; such behaviours are not necessarily verbal actions, and they can also be ‘communicative’ and sharing looks (Carpenter & Liebal, 2012) or gestures, such as pointing and showing (Siposova & Carpenter, 2019). To conclude, the lower attention levels (monitoring and common) include third-person perspectives; that is, the participants are *individually attending to the same thing*. The two higher attention levels (mutual and shared attention), in turn, include a second-person relation, which means that the participants are *jointly attending to the same thing*. (For an overview of the sliding scale of jointness, as Siposova and Carpenter [2019] call it, see Figure 1.)

![Figure 1](image)

*Figure 1. A scale of jointness in joint attention behaviour (JAB). Modified from Siposova and Carpenter (2019).*

2. **Operationalisation of Joint Attention Behaviour During Collaborative Problem Solving**

Basically, JAB can be divided into two types of behaviour—*initiating* joint attention and *responding* to joint attention (e.g. Mundy & Newell, 2007). In social interaction, *immediacy* is the key: normally, there is a tight timeframe within which communication occurs (or does not occur), such as a 2-second window, defined as a sequence consisting of recognisable initiating and responsive actions (Korkia kangas, 2018). Accordingly, in productive CPS, it is expected that a well-performing dyad
would organise their efforts in serial sequences of discussion (chat) and actions (manipulating artefacts), consisting of initiating and responsive actions with their partner. However, in CPS situations when students first tackle a new problem-solving task, they are not only expected to represent and refer to ‘relevant’ parts of the problem but also to somehow ensure that the partner understands what has been referred to (Zemel & Koschmann, 2013). Thus, the placement of gaze in relation to these other activities is an important measure to better understand the properties of JAB during CPS, especially how the ‘jointness’ and common knowledge is developed and materialised here.

3. Research Questions

In this paper, we ask the following questions:

1) How are ‘jointness’ and common knowledge in JAB materialised in dyadic interaction in a remote CPS context? What is the meaning of gaze in understanding JAB here?

2) Do the unique properties of remote dyadic interaction restrict/require JAB to arise, and are some ‘degrees’ of jointness more evident and valuable in CPS?

4. Methods

5.1 Dual-Space Interaction Environment as the Context of Study

The environment, as a ‘dual-space’ interaction space (Zemel & Koschmann, 2013), encompasses a chat property as a free form, synchronous interface, and a space with actionable artefacts that have either a symmetrical or asymmetrical outlook for the individuals. In a symmetrical task, stimulus content and actionable artefacts are equal for the partners, whereas in an asymmetrical task, the dyad is given a unique subset of resources for problem solving; alternatively, the screen view can be identical, but the permit to move certain objects or scroll the bars is divided between the partners. The success of one student depends on the behaviour of the other and the reactions offered (Care et al., 2016).

5.2 Participants, Study Setup, Data

The data were collected in a live eye-tracking situation (e.g. Korkiakangas, 2018; Dindar, Korkiakangas, Laitila, & Kärnä, 2017) from two student dyads from an initial teacher education programme in a Finnish university. In the study, dyads were physically located in separate cognitive labs. While completing the tasks in dyads, their eye movements were recorded with desktop eye trackers (screen-based; SensoMotoric Instruments [SMI] RED 250 Mobile), and a chin rest at a 60-cm viewing distance was used. A (13-point) calibration was conducted prior to the experiment and before each task. The dataset includes recorded observational data as activity logs from the online environment (consisting of time-stamped information of the movement of artefacts [i.e. actions], as well as the dyadic interactions via free-form chat interface; see Care, Griffin, Scoular, Awwal, & Zoanetti, 2015), combined with gaze data. In the paper, the focus is on a symmetrical task titled ‘Laughing Clowns’ (e.g. Care et al., 2015), in which, unknown to them, the students are presented each with a clown machine and 12 balls to be shared (see Figure 2 for a screen view of Student B). The screen views of Students A and B are mirror images of each other’s, where both can view which balls are being used by the partner but are unable to see how it is being used (i.e., the drop position of the ball in the clown’s mouth or the exit point when it comes out). In other words, the visual information that gets transmitted in real-time is the number of the balls used by partners and the location of the ball being used. However, the trajectory of the ball when in use by their partner is not visible to the other student. The students must place the balls into their clown’s mouth while the mouth is moving to determine the rule governing the direction the balls will go (entry: left, middle, right; and exit: positions 1, 2, 3). The goal is to determine whether their clown machines work in the same way. To do this, the dyad needs to share information and discuss the rules, negotiating how many balls they should each use. In this regard, communication via the chat interface is central to succeed in this task.
5.3 Data Analysis: Focussing on Event-Related Measures

In eye-tracking studies, the predominant focus has been on the overall looking times at predefined areas of interest (AOIs; ‘where’ questions; Falké-Ytter, Bölte, & Gredebäck, 2013). Yet, when studying social interaction, this may not be sufficient (Falké-Ytter et al., 2013). Combining ‘where’ people look with ‘when’ they look at the AOIs, that is, the timing of gazing, is critical to understand JAB as sequential interaction at the pair level during CPS processes. These event-related gaze measures, focussing on the interactional organisation of gaze, are more informative about what makes some instances of gazing ‘social’ (e.g. Dindar et al., 2017; Korkia-Kangas, 2018). In the current study, with the challenging dynamic scene of the dual-space remote environment, there were multiple behaviours of interest linked to JAB, such as gazing at the chat window, the actionable artefacts and the instructions. For analysing the gaze data, video exports with fixations and scan path data views using BeGaze software were produced. Fixation means maintaining the gaze on a single location, whereas a scan path data view shows gaze positions and eye events plotted on the stimulus video. (For an example of a video export as a scan path view and the AOIs of the symmetrical Laughing Clowns task, see Figure 2.)

![Figure 2](image)

Figure 2. Image from a video export (scan path data view) of the Laughing Clowns task with areas of interest (AOIs).

Note. The example of the AOIs include the following: 1) the instructions; 2) the chat property; 3) actionable artefacts, such as (a) the shared balls, (b) the clown’s head/mouth and (c) the ‘issue’ of the dropped balls as letter-number combinations; and 4) the solution.

To better understand JAB in terms of the typology of ‘jointness’ (Carpenter & Siposova, 2019) and to search for related behaviour during CPS, a manual qualitative coding procedure was applied. First, the focus was on the pair-level activity log data to code for initiating and responsive activities (i.e. chat and actions of Student A and B) during the CPS process. The aim was to search for meaningful interaction events in dyadic interaction with regard to JAB in CPS. Second, to better understand the triadic interaction (i.e. relationship between self, other and an object or their attention) during problem solving, these interaction events were further analysed in parallel with the eye-tracking video exports that show the fixations and scan path data views of the individuals while completing he tasks. These videos make visible the location and the order of the gaze cursor at specific AOIs during these selected events. At this point, the levels related to the spectrum of ‘jointness’, described as ‘a typology of social attention and social knowledge’ by Siposova and Carpenter (2019, p. 261), were identified from the data. It was assumed that these two data types (log files, gaze data), when analysed for consistency, would allow for better understanding of the interaction events in this regard. In addition, it was assumed that, in the remote environment, gaze data would show important moment(s) related to JAB during CPS, composed without writing or moving artefacts, comprising the following: (a) what the log file can show and (b) how the gaze data may reinforce or modify the initial interpretation of the interaction, based on the log file only. (For the different phases of analysis, see Table 1.)
Table 1. *The Different Phases of Analysis of the Event-Related Measures of Joint Attention Behaviour (JAB) During Collaborative Problem Solving*

<table>
<thead>
<tr>
<th>Phase of the analysis</th>
<th>Data source/level</th>
<th>Manual coding procedure</th>
<th>Target, expected output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 1</td>
<td>Pair-level activity log data</td>
<td>To code the initiating and responsive actions on the activity log file</td>
<td>To search for sequential characteristics of the log data (chat and actions) as meaningful events related to joint attention behaviour (JAB) in collaborative problem solving (CPS)</td>
</tr>
<tr>
<td>Phase 2</td>
<td>Gaze data, individual level (Students A and B), meaningful events (pair level) related to JAB in CPS, chosen at the previous phase</td>
<td>To code the location and order of the gaze cursor at specific areas of interest (AOIs) during the selected events (to analyse for congruence; frame-by-frame analysis)</td>
<td>To identify whether and how data reinforce or modify the initial interpretation of the interaction, based on the log file only, to better understand the interactional event at the pair level</td>
</tr>
</tbody>
</table>

*Note.* The table includes the different phases of the analysis, description of the manual coding procedure and the target or expected outcome(s).

5. Results

The first phase of the analysis provided a general overview of the sequential characteristics of dyadic interaction (i.e. the individual partner’s initiating and responding actions; actions, e.g. as the manipulations of the artefacts) and gave evidence of diverse orientations of the two dyads when completing the following tasks: (a) synchronised activity (Pair 1) and (b) parallel actions (Pair 2) (e.g. Pöysä-Tarhonen, Care, Awwal, & Häkkinen, 2017, 2018). In short, when focussing on the log stream of Pair 1, their interaction was well coordinated (Miles, Lumsden, Flannigan, Allsup, & Marie, 2017), and the contents and the general interactional organisation reflected the designed, task-specific CPS elements of the Laughing Clowns task, such as interaction, audience awareness, reciprocity and systematicity (see Care et al., 2015). In turn, when focussing on the log stream of Pair 2, the contents and the general interactional organisation lacked evidence of the majority of the designed, task-specific CPS elements. Dyadic interaction was parallel, encompassing autonomous actions of individuals, without systematicity or coordination, or as ‘trial–error’ actions (e.g. Davis et al., 2015). At the second phase, the identified patterns were analysed in greater depth with the gaze data, and resulted in illustrative examples that exemplify the spectrum of jointness (Siposova & Carpenter, 2019) in CPS, presented as follows: from individual and monitoring attention (Figure 3) to common attention (Figure 4); and from mutual attention to shared attention (Figure 5). The examples (see Figures 3–5) rely on both gaze and log stream data and are derived from two dyads accomplishing the task.

6.1 Individual and Monitoring Attention During Collaborative Problem Solving

Figure 3 exemplifies an individual (Student B) monitoring attention in the CPS situation, observed from the onset of the Laughing Clowns task. As typical of a monitoring situation (Siposova & Carpenter, 2019), Student B has individual knowledge of the situation and evidence of the partner, and via the screen, Student B attends to what the partner (Student A) is attending to. While dragging and dropping a ball, Student B takes an observer’s perspective on the actions of Student A; there are no communications yet, but Student B frequently monitors the screen and the interaction property. In this case, Student A is simultaneously in individual attention (Siposova & Carpenter, 2019), concentrating
on reviewing the instructions and testing the machine individually without any monitoring or communication via the chat property.

(a) Log stream data view (a dyad level):

<table>
<thead>
<tr>
<th>Action</th>
<th>Event Details</th>
<th>Start Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>B action</td>
<td>startDrag:ball1:76:150</td>
<td>6.3.2019 23:34</td>
</tr>
</tbody>
</table>

(b) Gaze data view (an individual level, Student B):

(c) Gaze data view (an individual level, Student A):

Figure 3. Illustration of a monitoring attention experience (Student B perspective) and individual attention experience (Student A perspective) during completing the ‘Laughing Clowns’ task.

Note. The example includes simultaneous moment from (a) the log data, combined with (b) and (c) individual level screen captures from the eye-tracking video exports.

6.2 Common Attention During Collaborative Problem Solving

As in monitoring attention, in common attention (see Figure 4), the experience is primarily individual (Siposova & Carpenter, 2019). Although working in parallel during CPS task completion and without systematically communicating over their related goals, Student A and B have the following characteristics: (a) they have an established joint objective, acquired via task instructions, and thus, it can be assumed that (b) their attention is relevant to the partner (Siposova & Carpenter, 2019). Here, the dyad is engaging in the same CPS situation; while they depend on the attention of the partner, their evaluation of the situation (common attention or not) is individual. In the log stream, the partners share their first notions as follows: Student B writes, ‘The first ball went into L’, and continues to test another ball without any further negotiation; immediately afterward, Student A writes, ‘The same and the head was left’. The communication is based on reporting parallel efforts that rely on individual partners testing the task-specific properties.

(a) Log stream data view (a dyad level):

<table>
<thead>
<tr>
<th>Action</th>
<th>Event Details</th>
<th>Start Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>B chat</td>
<td>The first ball went into L</td>
<td>6.3.2019 23:34</td>
</tr>
<tr>
<td>A chat</td>
<td>The same and the head was on left</td>
<td>6.3.2019 23:34</td>
</tr>
</tbody>
</table>

(b) Gaze data view (an individual level, Student B):

(c) Gaze data view (an individual level, Student A):

Figure 4. Illustration of a common attention experience (Student B and A perspectives).

Note. The example includes simultaneous moment from (a) the log data, combined with (b) and (c) individual level screen captures from the eye-tracking video exports.
6.3 From Mutual Attention to Shared Attention During Collaborative Problem Solving

As in the third-person perspectives of individual and common attention, two individuals ‘meet in the middle’ (Siposova & Carpenter, 2019, p. 262; see also Carpenter & Liebal, 2011), in the second-person relations of mutual and shared attention a ‘meeting of minds’ occurs, and the partners are truly attending with each other towards the shared goal or object of attention (Siposova & Carpenter, 2019). In the example (see Figure 5), at the onset of the task, the partner’s presence is verbally acknowledged—an attention contact is made. This can also be referred to here as a sign of mutual attention experience: In the remote environment, only verbal signs are available; the eye contacts typical of mutual attention in a physical environment are not possible. When proceeding with the task, both students’ communicative exchanges about the task and the task properties are intentional and bidirectional. In addition, over the course of CPS (except the first ball thrown by Student B without first consulting the partner, but Student B coming back to the issue later; see Figure 5), the partners co-create their experiences by constantly sending and receiving information and negotiating how to solve the problem together: The partners are seemingly engaged in ‘doing together’ as shared attention experience (Siposova & Carpenter, 2019; Zahavi, 2015). (The gaze data examples represented in Figure 5 are from the moment when the dyad explores the available artefacts and communicates on whether they both have two rows of six balls.)

(a) Log stream data view (a dyad level):

(b) Gaze data view (an individual level, Student A):

(c) Gaze data view (an individual level, Student B):

Figure 5. Illustration of sliding from a mutual to shared attention experience (Student A and B perspectives).

Note. The example includes simultaneous moment from (a) the log data, combined with (b) and (c) individual level screen captures from the eye-tracking video exports.

6. Discussion and Conclusions

The results gave empirical evidence of the theorised, third-person, individualistic attention experience (monitoring and common attention), as well as second-person relations (mutual and shared attention; see Siposova & Carpenter, 2019), in which both Students A and B adopted an engaged approach toward each other to solve the task together. In addition, it was observed how a lower level attention experience in CPS can be a precursor to a higher level of attention, such as moving from monitoring the partner’s actions towards common attention experience between partners. Furthermore, it was noticed that richer,
second-person relations may come in degrees. Siposova and Carpenter (2019) point out that generating shifts in the scale of jointness requires salient stimulus (here, the moving head of the clown), a relevant shared goal between the partners (to solve the problem of whether their machines work similarly) and a common ground (previous experiences and knowledge of the partner as a co-student). Moreover, the short time interval (immediate nature of sharing the appearance of the object and communication of the objects) and the limited perceptual space of the structured CPS environment can ‘push’ forward the situation in the scale of jointness. In our study, the two dyads were also rather distinctive in their orientations: The second dyad was seemingly more motivated to share ideas and thoughts than the first, which may reflect individual differences in this regard, helping to slide right on the scale of jointness (Siposova & Carpenter, 2019). Individual commitments can trigger social obligations beyond one’s own, and through communication, create joint goals and joint commitments to it (Siposova & Carpenter, 2019; Siposova, Tomasello, & Carpenter, 2018). This issue is closely connected to the quality of the behaviours—for example, how explicit or detailed the communication is. In our cases, there were also evident differences in this regard. Moreover, to fully understand social connotations in CPS, gaze data did not provide sufficient contextual details of the actual interactions alone, but, in the lower attention levels such as monitoring attention conditions, for example, the gaze data view was beneficial. This is because it not only reinforced the individual orientation, visible in the log stream data view, but also gave evidence of the ‘social’ behind these data. Due to the unique properties of remote social interaction, the ‘richness’ of attention experience or the indicative behaviours may not be fully identical if compared with a face-to-face situation as defined in Siposova & Carpenter (2019). Also, the attention experience and the indicative behaviours of JAB in remote dyadic interaction may not be fully equal with younger population of students, the authentic target audience of the designed CPS tasks (11-15-year-olds). As limitations of this study also include the low number of cases, as a next step, the focus is on larger population of middle school students’ CPS processes (40 dyads), and, particularly, on longer behavioural sequences of CPS interaction in dyads. That is, to search for more evidence by analysing what behaviours preceded and what followed after identified social attention levels in longer tasks with multiple student dyads in school context (see Siposova & Carpenter, 2019).

Acknowledgements

This research is funded by the Academy of Finland (Grant no. 316836).

References


Abstract: This study tests the proposal that a learner’s learning state can be estimated by observing their facial expressions. This could prove helpful for adaptive learning applications using facial recognition technology to provide appropriate feedback to learners. Based on our knowledge of facial expressions and use of the Interactive, Constructive, and Active processes from the ICAP framework, we hypothesized the following: (1) During the Active process, because utterances consist of reading the learning material, the muscles of the mouth are primarily in motion. (2) During the Constructive process, the muscles of the mouth move, and the eyelids tighten, because self-reflection and thoughtful utterances are required. (3) During the Interactive process, the conversation involves conflict; therefore, the eyebrows raise and the eyes open. To that end, we recorded and analyzed the facial expressions and utterances of five pairs of learners during the three learning states. We then organized the data by facial muscles that appeared most frequently during the respective learning states. The analyses generally supported our hypotheses. However, micro-facial expressions generated by facial muscles other than those considered in our hypotheses are also relevant and should be explored in future research.

Keywords: Facial expression analysis, ICAP framework, collaborative learning, learning support, adaptive learning

1. Introduction

Studies have found that during collaboration, constructive interaction that reconstructs one's understanding from different perspectives is important (Chi, De Leeuw, Chiu & Lavancher, 1994; Okada & Simon, 1997; Shirouzu, Miyake & Masukawa, 2002). In recent years, automatic estimation of the collaborative learning process has attracted attention in the field of educational technology. During collaborative learning, it is important to support learners based on their state of learning (Hayashi, 2019a). In this study, we focus on the learner's facial expressions during collaborative learning and examine whether they are effective indicators for estimating a learner’s state.

In Section 1.1, we explain the requirements for collaborative learning support. In Section 1.2, we discuss the research on learning support using facial expressions and explain the relationship between facial expressions and the collaborative process, which is the focus of this study. Then, we state our purposes and hypotheses (Section 1.3).

1.1 Support in collaborative learning

Researchers have confirmed that learning and performance are both promoted by interacting with others in cooperation (Chi, De Leeuw, Chiu & Lavancher, 1994; Okada & Simon, 1997; Shirouzu, Miyake & Masukawa, 2002). During tasks, learning in pairs is better than individual learning for making new discoveries, facilitating explanatory activities, and participating actively, resulting in high performance (Okada & Simon, 1997). Additionally, self-explanation during collaborative problem solving integrates existing knowledge with new knowledge (Chi, De Leeuw, Chiu & Lavancher, 1994). Furthermore, constructive interaction that reconstructs one's understanding based on different viewpoints in
collaborative learning is important in certain tasks, such as origami (Shirouzu, Miyake & Masukawa, 2002). Jigsaw learning research is an example of practical research that focuses on cooperative activities based on different perspectives. With the jigsaw learning method, students learn by bringing complementary knowledge to help solve a problem or complete a task. For example, Zhu & Sunakawa (2019) analyzed the interview data of students who took a class taught by the jigsaw learning method and examined the factors that caused many students to change from the "self-study type" to the "collaboration type." As an example of laboratory research, Hayashi, Miwa & Morita (2007) examined cooperative learning based on different viewpoints using an experimental task that uses the principle of illusion, but found that it was difficult to obtain the viewpoint of both learners in the dyad.

As described above, it is important to reconstruct one's understanding based on the different viewpoints of others. However, researchers have found that it is difficult to eliminate communication discrepancies during such learning activities, and that they can result in failure to acquire correct perspectives of others and constructive interaction (Hayashi, Miwa & Morita, 2007). Therefore, when creating a support system for collaborative learning, it is necessary to observe the collaborative process between learners and support it.

One of the tasks involved in developing a support system for collaborative learning is providing relevant feedback based on the state of the learner. This type of support is called adaptive feedback. An example of an adaptive feedback system is the Cognitive Tutor (Anderson, Corbett, Koedinger & Pelletier, 1995), which detects the state of the learner using a model and then provides relevant feedback. However, few studies have examined adaptive feedback in collaborative learning. Detecting learner states during the collaborative learning process, and providing them with appropriate feedback during such states, are research themes that are worthy of attention. For example, in a study using pedagogical conversation agents (like online assistants) to facilitate collaborative learning, Hayashi (2019b) established the effectiveness of estimating the collaborative process based on two indicators, gaze recurrence and language alignment. However, there is need to increase the number of indicators for detecting the state of the learner during collaborative learning. Therefore, in this study, we focus on facial expressions and non-verbal information, which are important during communication.

1.2 Detecting learner state using facial expression analysis

Facial expression research has been conducted for many years, and it is historically known that facial expressions can be used to estimate emotions (Ekman & Ancoli, 1980). The Facial Action Coding System (FACS) is a comprehensive system for distinguishing facial movements Ekman & Freisen (1976) using facial expression muscles, described as action units (AU). For example, the emotion worry is related to AU04 (Brow Lowerer) and AU07 (Lid Tightener). In another study using FACS, the relationship between the learner's facial expressions and emotional state was examined: facial expressions could help estimate emotions related to mutual understanding during collaborative learning (Hayashi, 2019b). These studies suggest that the learner state can be detected from the learner's facial expression. However, although we have investigated relationships between facial expressions and emotional states, it is necessary to investigate the relationship between facial expressions and the learning process from the context of providing adaptive feedback based on the learner's state.

The ICAP framework classifies learning activities into Passive, Active, Constructive, and Interactive (Chi & Wylie, 2014; Chi, 2009). In the present study, we examine the learner's activity state (Active, Constructive, and Interactive) during collaborative learning to capture the deep interaction of the collaboration (Chi & Wylie, 2014). The relationship between facial expressions and emotional states is not analyzed here.

Now let us look at the ICAP modes in detail. In the passive mode, it can be assumed that the knowledge change process is isolated in an encapsulated manner to "store" the received information. A learner directly collects information from learning materials during lessons, and there is no other behavior related to learning; for example, listening to a lecture without taking notes. In the active mode, the learner is not required to act, such as when reading learning materials. Once the relevant schema is activated, the learner can absorb or integrate new information into the activated schema, so that the learner can fill in the gaps in the model and make it more complete. In the constructive mode, the learner delves into the learning material and externalizes the content. Constructive behaviors often require the processes of "inferring" and inferring means a process of elaborating, such as adding more details or qualifications. For example, if you learn A and B and associate them, you then possess an utterance
about their contents and have created a link on the concept map. In the Interactive mode, a learner constructs thoughts based on what someone else has said; having doubts about the content; objecting to the content; or having a conversation involving conflicts. Each member of the dyad must be constructive, thereby engaging in the cognitive processes of activating, integrating, and inferring. For example, a learner finds a contradiction in the concept that the other learner explained and requests them to explain it further. Based on the above, when FACS is used to predict the learner's ICAP modes, the following will be true. In the Active mode, since the utterance is simply a reading of the learning material, it is mainly the muscles of the mouth which move. In the Constructive mode, since the utterance is engaged in self-reflection and results from thinking deeply, in addition to the above described movement of the muscles of the mouth, the eyelids tighten. In the Interactive mode, since the utterance is in a conversation involving conflicts, other facial muscles come into play and movements such as raising of the eyebrows and opening of the eyes are observable.

1.3 Purpose and hypothesis

To support the learner during collaborative learning, it is necessary to make a human facial expression model that can estimate the state of the learner during the collaborative process. Therefore, this study has two purposes. The first is to estimate which ICAP learning process a learner is engaged in based on their AUs. The second is to sort the types of AUs present in each ICAP modes. We will verify the following hypotheses.

H1-1: In the Active mode, since the utterance is simply a reading of the learning material, the main facial movement is from the mouth muscles, such as those are used to making dimples.
H1-2: In the Constructive mode, since the utterance is a result of self-reflection and thinking deeply, in addition to the muscles of the mouth, the eyelids tighten.
H1-3: In the Interactive mode, since the utterance is produced in a conversation involving conflicts, we can observe the movement of facial muscles such as those needed for raising the eyebrows and opening the eyes.

2. Method

The present study used the ten people from Shimojo & Hayashi (submitted) as a sample. Here, pairs of learners worked on a collaborative learning task that required explaining a specific psychological phenomenon while creating a concept map. For details, refer to (Shimojo & Hayashi, 2019a; Shimojo & Hayashi, 2019b; Shimojo & Hayashi, 2019c).

2.1 Participants analyzed in this study

We analyzed ten Japanese university students (1 male and 9 female) which had successful collaborative process. Two were excluded because they did not meet the criteria for classification within the ICAP framework.

2.2 Materials and systems

Two PCs and two monitors were used by the participants, and two Sony HDRCX680 video recorders filmed their conversations and facial expressions. We also used Cmap Tools, a tool for creating concept maps (https://cmap.ihmc.us/). A monitor and video recorder were placed in front of each participant in a pair. They sat across to each other and a partition was put between them so they cannot see each other.

2.3 Procedure

In the experiment, the task was to make inferences about a certain psychological phenomenon. Before the task, participants read a text about causal attribution. Then, there was a story that participants participate in school counseling with Michael Peter and he is worried about the new semester, they need
to explain why the student is worried about the new semester based on the story by causal attribution. Furthermore, the concept map was used as a tool for explaining activities, and the facial expression data during the conversation were collected.

In this study, based on the data in Shimojo & Hayashi (submitted), we analyzed the videos of the participants during the Interactive, Constructive, and Active modes; we verified their learning state by their utterances. We analyzed the video using OpenFace software.

2.4 Dependent variable: facial expression analysis

The utterances of the participants were classified using the ICAP framework, and the facial movements they demonstrated were analyzed by OpenFace in 1/20 second units. OpenFace automatically calculated whether an AU appeared. The numerical value output by OpenFace indicated the strength of the AU and was obtained by a formula described in a “toolkit” for using the software (Baltrušaitis, Robinson & Morency, 2016). There were 18 types of AUs observed among the participants as follows. AU01: Inner Brow Raiser, AU02: Outer Brow Raiser, AU04: Brow Lowerer, AU05: Upper Lid Raiser, AU06: Cheek Raiser, AU07: Lid Tightener, AU09: Nose Wrinkler, AU10: Upper Lip Raiser, AU12: Lip Corner Puller, AU14: Dimpler, AU15: Lip Corner Depressor, AU17: Chin Raiser, AU20: Lip stretcher, AU23: Lip Tightener, AU25: Lips part, AU26: Jaw Drop, AU28: Lip Suck, and AU45: Blink.

After determining the AUs involved, we analyzed the video of the utterances for each ICAP learning process. The frequency of each AU was tabulated for each video (second). In the analysis, the Z score was calculated. The video comprises 418.72 minutes of the Interactive mode, 158.19 minutes of the Constructive mode, and 123.85 minutes of the Active mode.

3. Result

3.1 Checking the reliability of automatic coding

Before verifying the hypothesis, we confirmed the reliability of the AUs calculated by OpenFace. We analyzed 6 of the 18 AUs by sampling all participants’ data. Then, we set the coding standard for each AU facial expression, and performed manual labeling. For example, the coding standard for AU45 was defined as the movement of "closing and opening eyes."

To calculate the degree of agreement between the manual and the OpenFace coding results, we converted the data to obtain the Z score. We analyzed human detection of AUs against OpenFace detection of AUs. Table 1 demonstrates the overall concordance rate. From this result, Table 1 shows that the degree of agreement regarding AU07 (Lid Tightener) is low, but that the degree of agreement regarding the other AUs is high. The analysis revealed that there was the correlation of Z score between manual coding and OpenFace coding. Based on these results, we attempted to confirm H1-1, H1-2, and H1-3 using OpenFace coding.

Table 1. Concordance rate by AU for analysis by human and OpenFace

<table>
<thead>
<tr>
<th>Action Unit</th>
<th>Interactive</th>
<th>Constructive</th>
<th>Active</th>
</tr>
</thead>
<tbody>
<tr>
<td>AU02 (Outer Brow Raiser)</td>
<td>87%</td>
<td>90%</td>
<td>85%</td>
</tr>
<tr>
<td>AU05 (Upper Lid Raiser)</td>
<td>100%</td>
<td>55%</td>
<td>95%</td>
</tr>
<tr>
<td>AU07 (Lid Tightener)</td>
<td>12%</td>
<td>6%</td>
<td>29%</td>
</tr>
<tr>
<td>AU17 (Chin Raiser)</td>
<td>76%</td>
<td>99%</td>
<td>78%</td>
</tr>
<tr>
<td>AU20 (Lip stretcher)</td>
<td>98%</td>
<td>100%</td>
<td>95%</td>
</tr>
<tr>
<td>AU45 (Blink)</td>
<td>92%</td>
<td>95%</td>
<td>95%</td>
</tr>
</tbody>
</table>

3.2 Analysis of the relationship between ICAP and AUs

In this section, we describe the results of the OpenFace analysis and show which AUs appeared during each mode of the ICAP process. Figure 1 shows the Z score of each AU for each ICAP mode, and the details of the analysis results are described below.
3.2.1 AUs that frequently appear in the Active mode

In the Active mode, the frequency of AU05 (Upper Lid Raiser), AU14 (Dimpler), and AU45 (Blink) is high. The Z score was calculated, and a score higher than 1 was judged to be high in frequency. The same standard was applied to the Constructive and Interactive modes. We also examined the difference between AUs higher than 1. To examine which facial expression muscle appeared most strongly, we compared the average values of AU05 (Upper Lid Raiser), AU14 (Dimpler), and AU45 (Blink).
Therefore, a one-factor within-participant analysis of variance was performed and no significant difference was confirmed ($F(2, 14) = 1.80, p = .20, \eta^2_p = .21$). All three AUs appeared at the same level.

From the above, it can be said that H1-1 was generally supported because AU14(Dimpler) appeared at a high frequency during the Active mode.

### 3.2.2 AUs that frequently appear in the Constructive mode

In the Constructive mode, we found that the frequency of AU02 (Outer Brow Raiser), AU06 (Cheek Raiser), AU07 (Lid Tightener), AU09 (Nose Wrinkler), AU12 (Lip Corner Puller), AU15 (Lip Corner Depressor), AU23 (Lip Tightener) and AU25 (Lips part) was high. Then, the average values of AU02 (Outer Brow Raiser), AU06 (Cheek Raiser), AU07 (Lid Tightener), AU09 (Nose Wrinkler), AU12 (Lip Corner Puller), AU15 (Lip Corner Depressor), AU23 (Lip Tightener) and AU25 (Lips part) with Z score higher than 1 were compared. A one-factor experiment within-participant analysis of variance confirmed a significant difference ($F(7, 49) = 3.48, p < .01$). Therefore, when multiple comparisons were performed to confirm differences, no significant difference was observed. These AUs appeared to the same extent.

Based on the above, the frequency of AU07 (Lid Tightener) was high and the frequency of AU14 was not high, but the frequency of AU12 (Lip Corner Puller), AU15 (Lip Corner Depressor), AU23 (Lip Tightener) and AU25 (Lips part), which are the movements of the mouth muscles, were high.

### 3.2.3 AUs that frequently appear in the Interactive mode

By the same criteria used for the Active and Constructive modes, we found that AU01 (Inner Brow Raiser) and AU28 (Lip Suck) had a high frequency in the Interactive mode. Then, we compared the average values of AU01 (Inner Brow Raiser) and AU28 (Lip Suck) with Z scores higher than 1. A paired t-test revealed that AU01 (Inner Brow Raiser) was significantly higher than AU28 (Lip Suck) ($t(7) = 4.12, p < .01$).

Therefore, the frequency of AU05 (Upper Lid Raiser) is not high, but the frequency of AU01 (Inner Brow Raiser) is high. Therefore, it can be said that H1-3 was partially supported.

### 3.2.4 Concluding remarks on AUs that appear frequently

The statistical analysis clarified that different AUs appear in each mode of ICAP’s collaborative learning process. The study’s second purpose was to categorize AU types by their occurrence during each ICAP mode, as summarized in Table 2. The order of the type of the AU shown in the table occurrence frequency is placed in the order by the type of AU with high frequency.

From this we determined that there are no AUs that overlap the different ICAP modes; as such, this method is useful for estimating which ICAP mode the learner is experiencing. A more detailed discussion of this point will be provided in the upcoming section.

**Table 2. AUs that have a high frequency of occurrence by ICA**

<table>
<thead>
<tr>
<th>ICAP framework</th>
<th>AU that frequently appear</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interactive</td>
<td>AU01 (Inner Brow Raiser), AU28 (Lip Suck)</td>
</tr>
<tr>
<td>Constructive</td>
<td>AU02 (Outer Brow Raiser), AU06 (Cheek Raiser), AU07 (Lid Tightener), AU09 (Nose Wrinkler), AU12 (Lip Corner Puller), AU15 (Lip Corner Depressor), AU23 (Lip Tightener), AU25 (Lips part)</td>
</tr>
<tr>
<td>Active</td>
<td>AU05 (Upper Lid Raiser), AU14 (Dimpler), AU45 (Blink)</td>
</tr>
</tbody>
</table>
4. Discussion

In H1-1 we predicted that the muscles that move the mouth, such as when making dimples, were the main type of AUs present during the Active mode. In H1-2, we predicted that during the Constructive mode, in addition to the muscles of the moving mouth, the eyelids would tighten. In H1-3, we predicted frequent eyebrow raising and eye opening during the Interactive mode. Our examination revealed that the frequency of AU05 (Upper Lid Raiser), AU14 (Dimpler), and AU45 (Blink) was high during the Active mode; AU02 (Outer Brow Raiser), AU06 (Cheek Raiser), and AU07 (Lid Tightener) during the Constructive mode; and AU09 (Nose Wrinkler), AU12 (Lip Corner Puller), AU15 (Lip Corner Depressor), AU23 (Lip Tightener), AU25 (Lips part), AU01 (Inner Brow Raiser), and AU28 (Lip Suck) during the Interactive mode. H1-1 is generally supported by these results because of the high frequency of AU14 (Dimpler); H1-2 is generally supported because of the high frequency of AU07 (Lid Tightener), AU12 (Lip Corner Puller), AU15 (Lip Corner Depressor), AU23 (Lip Tightener) and AU25 (Lips part); and H1-3 is partially supported because of the high frequency of AU01 (Inner Brow Raiser). In addition, we found that different high frequency AUs appeared in each ICAP phase. As there is no overlap of high frequency AUs between phases, observing AUs can be considered useful for estimating ICAP. For example, if AU01 (Inner Brow Raiser) and AU28 (Lip Suck) both appeared on the face of the learner, there is a high possibility that the learner is in the Interactive mode. Furthermore, results show that micro-facial expressions produced by facial muscles that were not mentioned in the hypotheses were additionally found. For example, the frequency of AU09 (Nose Wrinkler) was a type of muscles that we did not consider in our hypotheses, but was detected during the Constructive mode.

Previous studies have attempted to estimate emotions by analyzing facial expressions. For example, researchers have clarified that facial expressions can be used to infer emotions related to mutual understanding in collaborative learning (Hayashi, 2019b). However, there has been little research on estimating the learner’s mode during the collaborative process, which is required for adaptive feedback. Moreover, this is the first attempt to classify facial expressions performed during the ICAP modes proposed in Chi and Wylie’s work on the topic; they categorized the ICAP mode of the learner by utterance and behavior (Chi & Wylie, 2014). Because the AUs with a high frequency of appearance differed for each ICAP mode, AU can estimate the learning process of ICA. Thereby, providing adaptive feedback based on the learner's learning states will be possible. So, this study provides important knowledge for effective adaptive support in future collaborative learning models.

In addition, ICAP is classified by utterances and actions (Chi & Wylie, 2014). In this study, we analyzed the data classified into ICAP based on the learner's conversational activities. Therefore, it is necessary to more strictly classify into ICAP by not only the conversational activities but also the indicators such as learner's behavior and gazes. ICAP predicts different learning levels, Active realizes a higher learning level than Passive, Constructive realizes a higher learning level than Active, and Interactive realizes a higher learning level than Constructive (Chi & Wylie, 2014). In this study, we examined AUs that have a high frequency of occurrence for each ICA, but we did not consider Passive. Since the learner's state needs to be promoted the transition from Passive to Active, it is necessary to provide adaptive feedback when detecting Passive. Therefore, it is considered that the passive feedback can be estimated from the AU and adaptive feedback can be performed by examining the AU that frequently appears in the passive.

5. Conclusion

In this study, we explored how learners can be better supported during collaborative learning by quantitatively investigating learners’ learning process. To that end, we analyzed facial expressions during different ICAP modes by observing the movement of their AUs during the aforementioned states. In the present study, (1) we found that the learning state can be estimated by observing the learner’s facial muscles, and (2) we organized the relevant AUs by their frequency during the respective ICAP modes. We used utterance and facial expression data obtained from pairs of learners while they performed a collaborative learning concept explanation task. The results of the data analysis generally supported our three hypotheses. On the other hand, it became clear that micro-facial expressions produced by facial muscles other than those considered in the hypotheses are also related. In future
studies based on the AUs classified in this study, we will examine their relationship with the learning process in more detail and perform deeper quantitative analysis to fine tune the development of effective adaptive feedback models.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP20H04299.

References


Learning Analytics of Critical Reading Activity: Reading *Hayavadana* during Lockdown

Rwitajit MAJUMDAR*a, Geetha BAKILAPADAVU*b, Reek MAJUMDER*c, Mei-Rong Alice CHEN*a, Brendan FLANAGAN*a & Hiroaki OGATA*a

aAcademic Center for Computing and Media Studies, Kyoto University, Japan
bDepartment of Humanities and Social Science, BITS Pilani Goa Campus, India
cSchool of Computing, Clemson University, USA
*dr.rwito@gmail.com

Abstract: Investigating learning behaviors in a humanities course using learning analytics techniques is underrepresented in literature. A *Critical Analysis of Literature and Cinema* course was selected as a context. The course was offered for more than 10 years by the instructor in a face-to-face classroom mode. However, this time the context was unique as the classroom-based interactive activities were rapidly migrated to online sessions due to the COVID-19 pandemic related lockdown. Under such a circumstance, BookRoll, a learning analytics enhanced eBook platform supported the critical reading activity online. Students (n=22 out of the 50 registered) accessed *Hayavadana*, an Indian play titled *Hayavadana* uploaded on BookRoll and attempted to identify performative elements and cultural references in the text and highlight them. In this study, we analyze learner’s reading logs gathered in the learning record store linked to BookRoll during that activity. Based on learner’s online reading engagement from their clickstream interactions and time spent for them, four readers’ profiles were defined; *Effortful, Strategic, Wanderers* and *Check-out*. We illustrate the content navigation and annotation behavior of each of those profiles. This study aims to initiate further discussion related to the application of learning analytics in humanities courses both to enhance the teaching and learning experiences by the use of interactive learning dashboards that was used to probe into the learning behaviors of the students.

Keywords: Learning Analytics, Humanities Course, BookRoll, Critical Reading Activity, Process Mining, *Hayavadana*

1. Background and Motivation

Learning analytics as a domain has evolved over the last decade to apply various computational techniques to collect, analyze and understand data related to teaching-learning experiences and thereby enhance them. While there are many studies which look at the different applications of Learning Analytics (LA) in STEM domains (Sergis et al. 2019), the focus on humanities courses at the university level is still limited. In this pilot attempt, we investigate student’s behaviors at a critical literature analysis course during their specific activity to analyze an act of a play. The play *Hayavadana* by Girish Karnad was chosen for the activity and it was done online using BookRoll, a Learning Analytics enhanced e-book-based learning platform. Learning traces are automatically collected as interaction logs based on the reader’s actions (navigation by clicking on buttons, annotating by using highlighting functions, etc) which could then be analyzed. Understanding the process of development of critical thinking skills is an important research aspect (Douglas, 2000). It can potentially help to inform teachers to design learning activities to support their students and also system developers to create technology assistance to orchestrate those activities.

Hence, in such an authentic natural learning setting, we investigate the following two research questions:

1. What are the reading behaviors of the learners given the critical reading task in terms of interactions within the e-content and the time spent on that task?
2. What are the different profiles of learners based on the distribution of the interactions and time spent with the e-content, and what are the characteristics in terms of content navigation and annotation patterns during the critical analysis task?

The article is organized in the following sections. Section 2 looks at the related works and provides the foundation of the study. Section 3 illustrates the context, learning task and the research methods. Section 4 presents the results of the analysis. Section 5 ends with the discussion and conclusion of the study.

2. Related work and foundation of the study

2.1 Critical reading activities

Critical reading is an active, in-depth reading of a text that calls forth a deeper engagement with the text. Such an activity requires cognitive tasks such as comprehending, analyzing, evaluating, interpreting and synthesizing. A critical reading activity requires one to highlight important ideas in the text, relate it to one’s personal experiences, pose questions and think about answers for such questions, look into the patterns within the text, and make connections with other texts. In certain contexts, it would involve identifying socio-cultural contexts and reading through them. Critical reading enables the reader to read not only the explicit meanings but the layered and the implicit meanings as well. Over and above, critical reading enhances one's ability for task-focused thinking. Critical thinking is reasonable, reflective thinking, focusing on a task, people or belief (Ennis, 1993). Also referred to as ‘good thinking’, ‘thinking well’ and ‘smart thinking’ - it enables one to identify questions worth pursuing through self-directed search and interrogation of knowledge (Pithers, 2000).

From the humanities education standpoint, developing critical reading skills is crucial. One of the essential values of Humanities is identified as critical thinking (Holm et.al. 2015). Especially so in the case of courses that deal with cultural texts including narrative arts. While understanding subjective experiences embedded in the texts and relating with them are important, deciphering the layered meanings is also equally significant. Developing critical reading skills enable one to do all the above tasks for a much-enriched meaning-making process.

In the context of reading a cultural text, it is imperative for the reader to be able to identify various pointers (references) that bring the cultural context to the fore. It is well established that the reader’s prior exposure to the cultural context enhances the reader's relatability to the text. The instructor, through her classroom experiences and reflections, had identified the role of annotation in making the reader dive deeper into the text. In this particular module being discussed, the text chosen is loaded in its cultural references and certain traditional theatrical conventions as performative elements. Hence, the instructor identified the learning task to annotate cultural references and performative elements. Learning logs were gathered for analytics during that task.

2.2 Learning analytics and critical reading activity

Critical reading from the perspective of critical thinking by the learners was studied using a technology environment at the school level in Singapore (Tan, Yang, Koh & Jonathan, 2016; Jonathan et al. 2017). The study used a collaborative environment and a LA dashboard for supporting critical reading activities at the school level. User acceptance of the innovation and its associated usability issues were its main focus. Our earlier studies with university level students proposed a specific in-class pedagogical model and studied their reading behaviors while they were comprehending English as a foreign language (Chen et al. 2019). Other works looked at embedding strategy prompts in digital text and found it had a positive effect on learners’ cognitive load, achievement, attitudes, metacomprehension and calibration accuracy (Reid, Morrison & Bol, 2017). However, another study from the perspective of the influence of the media (digital text vs physical text) on comprehension of text (Ben-Yehudah & Eshet-Alkalai, 2018) found learners improved only in printed condition while answering questions that required inferential processing.

Our study focused on gathering learning traces during an authentic learning task related to critical reading. Such a process-driven narrative regarding learner’s behavior in a critical reading task, specifically in a humanities course, is relatively rare and we want to fill in that knowledge gap.
2.3 Learning Evidence Analytics Framework (LEAF)

Learning Evidence Analytics Framework (LEAF) is an overarching technology framework to collect evidence of learning and teaching from the logs generated in a technology-enhanced learning environment (Ogata et al. 2018). In this instantiation of the framework, the instructor coordinated the course on Moodle, an LMS. BookRoll, an e-book reader, was used to upload reading contents like lecture slides, reference articles and reading assignments in PDF format for students to access. Tools like BookRoll can be considered as a learning behavior sensor as it can log student’s reading and annotation interactions in a Learning Record Store (LRS) as standard eXperience API (xAPI) statements. Figure 1 presents the technical architecture based on LEAF that is used in our study and the user’s reading interface in BookRoll which supports annotation functions such as highlighting with different colors, adding memos and bookmarks in the content. As long as there is an internet connection, students can read their books anytime from a web browser on their personal computer or smartphones. Student’s reading activity log from the LRS is then provided to the dashboard database and visualized for both the instructors and students appropriately.

3. Study Context and Method

As an initiative of a collaborative research project, access to the LEAF platform was given to the instructor and her students at one of the private universities in India. An overall phenomenographic research approach was chosen for the study (Marton 1981, Jan Larsson & Inger Holmström, 2007). It guided the research questions to focus on a single activity undertaken by the students enrolled in the course. The team of researchers including the course instructor then interpreted the different approaches that emerged from the learning logs.

3.1 Context

3.1.1 Course and its Objective

This particular study was conducted in an undergraduate elective course, Critical Analysis of Literature and Cinema (CALC) offered by the Humanities and Social Sciences Department. The objectives of the CALC course are three-fold: 1. to inculcate in students a critical insight required to interpret a work of literature and cinema, 2. to enable the students to perceive the subtle nuances of such works and to develop critical judgment, and 3. to introduce different forms, terminologies and trends prevalent in such artistic ventures to enable them to place a work of art in the proper context. The class was scheduled for 3 hours each week, split across three sessions. Students met for a total of fifteen weeks. In addition to these classroom interactions, students were given take-home readings and film viewings. The semester that just concluded had to undergo a sudden change of plan due to the pandemic and the early lockdown. It was the 9th week into the semester that the instructor had to shift the regular classes to online mode.
3.1.2 Participants

Students enrolled in the course (n=50, 17 males, 23 females) were pursuing their undergraduate program in engineering and sciences in the university. Their ages ranged from 19 to 23 years and the class included students in their second, third or fourth year of study in the university. At the time of the research, they had been introduced to approximately 1 to 3 humanities courses as electives. All the students were registered on Moodle for the following activity.

3.1.3 Critical reading activity: Instructions and example

One of the modules in the course related to critical reading and analysis of a play was orchestrated on the LEAF platform. The instructor chose an Indian play titled Hayavadana (Karnad, 1972), originally written in Kannada and then translated into English by the playwright himself. The content was uploaded on BookRoll and the students were given the task of going through the first act of the play to identify and highlight the following 1. cultural references (red highlight) and 2. performative elements (yellow highlight) in the text designated. The activity was designed around these two factors as the play is deeply steeped in the cultural milieu of traditional Indian theatre. Also, these two tasks were significant for a critical understanding of the play, however it was not a graded activity in the course.

Instructions for the task given to the students were posted on the LMS followed by announcing them to do it during the online synchronous class. There was no intervention from the instructor’s side during the reading and annotation activity. An example of the cultural reference and the performative element is shown below. We have selected 2 of the pages which the students have spent most of the time (see section 4.1)

![Figure 2. Cultural references and performative elements in page 8 and page 12 of content.](image)

3.2 Data Collection and Analysis

The data extracted for this study included the reading logs from 13 April 2020 to 1 May 2020 in BookRoll. Of the total 50 students enrolled in this course, 22 accessed BookRoll for the reading activity. 2587 logs were captured when the students were reading the play. For this study, we considered the 1575 action logs of opening the play, navigating through its pages and annotating its content. The highlighting event is considered towards fulfilling the given critical reading task.

To answer RQ1, data was collected from the learning analytics dashboard regarding the students access count, their annotations and the time spent on the specific page content. Using the data export option in the dashboard the interaction data for each student was retrieved for further analysis. We consider the following interactions OPEN, ADD MARKER, ADD MEMO, NEXT and PREV from the LAViEW database.

To answer RQ2, we defined four basic profiles of the readers based on their interaction log counts and the interaction duration in the BookRoll system. They are Effortful, Strategic, Wanderers and Check-outs (see Table 1 for definition). The count and duration are considered as a spectrum and for the pilot study we categorized the members based on the actual dataset. Then data is then processed for one member of each profile to illustrate the characteristics of that profile in terms of time spent, navigation pattern and highlighting activity in page 8 and 12 of the uploaded content.
Table 1. Initial profiles of readers based on their interaction log counts and interaction duration.

<table>
<thead>
<tr>
<th>Low Interaction counts</th>
<th>High Interaction counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Interaction Duration</td>
<td>Low Interaction Duration</td>
</tr>
<tr>
<td>Strategic</td>
<td>Effortful</td>
</tr>
<tr>
<td>Readers who focus on attempting the task and hence activity is concentrated on specific pages related to the task.</td>
<td>Readers browse through the content and may revisit and go back and forth multiple times while attempting the task.</td>
</tr>
<tr>
<td>Check-out</td>
<td>Wanderer</td>
</tr>
<tr>
<td>Readers who just opened the content and then left.</td>
<td>Readers who just navigated the content without any particular focus on the task.</td>
</tr>
</tbody>
</table>

4. Results and Interpretations

4.1 Information Available in the Dashboard: Overview Statistics and Time Spent on a Page

Figure 3 shows the overview of the content interaction information as seen in LAViEW, the learning analytics dashboard. The first column of information has the total learners who accessed the material and the total number of pages in the content. When one of the students is selected, it shows the average value of the student on the top and the class average on the bottom in the engagement section and the reading activity section. Here student #2548 is selected as an example case. It shows that the student kept the content open for 108.2 minutes when the class average recorded was 35.5 minutes and completion of the material as measured by pages browsed in 83% compared to the class average of 41.4%. The student had a total of 326 events with 47.2% of them more than 3 seconds duration (considered as a long event). The count of the reading activities such as maker and memo annotations are also given in the overview widget for both the teacher and the students.

The aggregated average page-wise viewing time is presented in Figure 4. It is calculated by considering content open as a proxy. The difference in time from landing on a page till the next-page or close interaction is considered as time spent on a particular page. Average page-wise viewing time was 2 minutes considering all the users who viewed the book. Based on this distribution, for this content we analyzed annotation activities only in the pages with the top two average times spent, page 8 (4.3 minutes) and page 12 (4 minutes).
4.2 Computed Learner-wise details

The distribution of the number of interactions and time spent on each interaction aggregated across all their reading sessions for each of the learners are presented in Figure 6. The figure also indicates the members of different reading profiles and presents the average characteristic values of that profile.

![Figure 6. Learner-wise profile based on their total interactions with the content and time spent.](image)
4.3 Characteristics of the Profiles of Critical Reading Activity

Next to answer RQ2, we present the details of candidate members of each profile.

4.3.1 Effortful Reader

Learner 2548 is selected to illustrate an Effortful Reader. The average page viewing time was 3 minutes (Fig 7a). The navigation pattern is presented in Figure 7b. The blue lines indicate the learner did NEXT on a specific page, and the red indicates PREV. The y-axis has the accumulated count of those transitions. Highlighted portions on page 8 and 12 by the learner are shown in Figure 7c and 7d.

4.3.2 Strategic Reader

Learner 2546 is selected to illustrate a Strategic learner. The strategic learner focused on the area of the task (Fig 8b) and spent more time (average 4 mins) across the pages (Fig 8a). The specific highlighted portions on page 8 and 12 are shown in Figure 8c and 8d.
4.3.3 Wanderer and Check-out

Learner 2557 is designated as a Wanderer reader who navigated through the content (Fig 9b) and spent time (average 4 mins) but did not attend to the task. The navigation pattern is presented in Figure 9b.

5. Discussions and Conclusions

5.1.1 Findings and contributions

This study investigated the reading behaviors of learners during critical reading tasks executed on an online e-book reader. This learning logs of interactions in the e-book system and the processed data from the learning analytics dashboard were used to define and describe four different reader’s profiles: Effortful, Strategic, Wanderer and Check-outs. While Effortful and Strategic readers attempted to complete the critical reading based highlighting task, the Wanderer and Checkouts didn’t attempt it at all. Visualization of the navigation patterns of learners in an e-book context was presented in section 4. It is used to illustrate the characteristics of different profiles. This is also the first attempt to discuss the quantitative account of reading the specific play, Hayavadana.

5.1.2 Critical analysis of Hayavadana: reflections from data

The play Hayavadana is firmly rooted in the Indian tradition and Hindu mythology and yet uses certain subversive tactics to question some of these deep-seated traditions and beliefs. At the performative level, it draws from Yakshagana, a traditional theatrical form practised in the coastal belt of Karnata, a state in southern India. It also uses a narrator, a chorus and a self-reflexive plot within plot strategy thereby bringing in much talked about alienation effect. Before any critical analysis of the play can commence, students are nudged to engage in a critical reading task- which typically is done before the...
classroom discussions. In this study, the activity was done fully using an online environment. The log data provides reading and annotation behavior of a group of learners interacting with the content of the play for the first time. Even though the readers were uninitiated about the conventions of Yakshagana, from the highlights drawn during their reading, it is evident that they did identify simple directions for performance. Tagging such simple directorial notes as performative elements can be attributed to their unfamiliarity with the more nuanced elements taken from the Yakshagana style. Further, the navigation graph data is in congruence with the above reasoning. It shows that Effortful and Strategic readers are particularly engaged in these parts of the text. The data also suggests these readers were clearer and more accurate in the tagging of cultural references as compared to performative elements. Such data-points give a clear indication to the instructor as to where the focus of discussions should be when she goes to the class - whether face to face or online.

### 5.1.3 Developing profiles of reflective reading and implications for technology design

In earlier works, Binder and Lee (2012) proposed four types of adult readers: Unskilled Readers, Resilient Readers, Good Decoders/Poor Comprehenders and Skilled Reader. Later, Putro & Lee (2018) conducted a latent profile analysis of readers across different modes (printed, online, and social media) and for different purposes (academic and recreational) of reading. They classified low-interest readers, traditional readers, moderate readers and high-interest readers. Still, specifically for critical reading, previous literature lacks any reader’s profile. We attempted to approach and fill that gap using learning logs and computing broader navigation patterns of different readers.

Reading strategies and comprehension strategies are considered as cognitive action and remedial action respectively and both assists the learners in achieving reading success (Yang, 2006). A technology framework like LEAF is capable of supporting these aspects by collecting learning logs from the e-reader and using learning dashboards to visualize the traces. Recent work (Gibson et al. 2017) focused on data-driven technology-supported feedback for reflective writing. However, for reflective reading activities, such data-informed digital services are still lacking. This study conceptualized using the interaction count and time as indicators of different profiles of readers. Such indicators are often included in LA dashboards (Tan et al. 2016) and can assist the teachers to directly check the visualized data and decide the status of reflective reading behavior of the learner.

At another level, technical support can also be developed to automatically evaluate the highlighting actions of learners and to give them feedback. During the data analysis process, the instructor highlighted the portions of the text for reference. Presenting the instructor’s highlighted part to the learners in the learning dashboard can also assist their learners.

### 5.1.4 Limitations and future work

This was a pilot attempt to understand and share some of the observed reading patterns and discuss possible ways the learner interacted with the task at hand. Some of the collected data remains difficult to interpret, for instance, the Wanderers, who spent time within the content without attempting the task (indicated by annotation action) it is not possible to distinguish whether they are coping up with comprehending the text before engaging in the critical analysis or just being off task in the system. While consolidating the action logs of 22 learners (44% of the registered participants) generated our dataset, it is still from a smaller sample space to fully comment on critical reading behaviors. This might be primarily due to the fact that the activity was ungraded. Further, in this analysis, we did not consider any explicit learner output apart from the highlights as the artefacts. A model without actual learners’ output and only having clickstream interactions has its own limitations regarding validity of the profiles generated. Hence given our analysis, we claim that the profiles are only of the readers based on the engagement attributes (time and clicks), and cannot distinguish learners yet. As a future work there remains further analysis of the data from the pilot study itself. We aim to investigate the quality of the highlighted text by the learner with respect to the instructor’s annotation and further compute inferential statistics for the difference of the profiles identified. These would lead to developing learner models specific to critical reading activities.
Acknowledgements

This research was supported by JSPS KAKENHI 16H06304, 18H05746 (F.Y.2018), 19K20942 (F.Y.2019), 20K20131, 20H01722, SPIRITS 2020 of Kyoto University and Teaching Learning Centre (TLC), BITS Pilani K. K. Birla Goa Campus (2019).

References


Using Log and Discourse Analysis to Improve Understanding of Collaborative Programming

Bernard YETT*, Caitlin SNYDERa, Ningyu ZHANGa, Nicole HUTCHINSa, Shitanshu MISHRAa & Gautam BISWASa

*aVanderbilt University, USA
*bernard.h.yett@vanderbilt.edu

Abstract: The importance and ubiquity of computing and computational thinking (CT) is leading to re-design of K-12 curricula and the development of appropriate platforms to support computer science instruction. One such environment is the block-based, synchronous programming environment, NetsBlox, that allows for real-time collaboration amongst students. This work presents a novel combination of discourse and activity log analyses to study the collaborative behaviors of K-12 students as they worked on a week-long cybersecurity curriculum. Groups of students were assessed based on pre-post-test learning gains in cybersecurity and CT. We analyze the differences between the collaborative behaviors and discourse of high and low performing groups using case study and differential sequence mining analyses to characterize productive and unproductive collaborative problem solving in programming tasks.

Keywords: computer science education, data-driven analysis, collaboration, robotics

1. Introduction

Collaboration is considered a key computational practice (College Board, 2020) and offers several benefits worthy of further exploration. Previous work has shown that collaboration provides social support, thus allowing a group to learn from each other and overcome difficulties (Saenz-Otero et al., 2010). Werner et al. (2012) found a significant positive relationship between time spent programming with a partner and assessment scores. Despite these, there is evidence of problems that can arise while students work together on programming projects. When pairing students with different experience levels, the more knowledgeable student often assumes full control and leaves their partner confused and disengaged (Braught, MacCormick, & Wahls, 2010). Working with others can be a frustrating, challenging experience that may ultimately undermine motivation (Rogat, Linnenbrink-Garcia, & DiDonato, 2013).

Taken together, it becomes important to design environments that guide students towards the benefits of collaboration while avoiding the pitfalls. To accomplish this, we need an improved understanding of the problem-solving processes groups employ, specifically the interweaving of dialog and actions as they work on their programming tasks. This may reveal how students develop an understanding of the desired computer science concepts and practices, while also exposing their difficulties and how they may overcome them via collaboration. Each mode of interaction requires unique analysis techniques in order to uncover the collaborative processes at play and ensure the success of each group. Our specific approach to collaborative programming analysis integrates discourse, learning, and log data, a combination that is still under-researched.

Cybersecurity was the primary focus of this intervention. It has become an important component of computer science (CS) curricula, but presents complex concepts and can be difficult to grasp when not situated within meaningful contexts (Thompson et al., 2018). Robotics platforms have been successful in delivering cybersecurity curricula to K-12 students (Karaman et al., 2017; Lédeczi et al., 2019; Yett et al., 2020a). These platforms allow students to collaborate to solve larger, more complex problems through co-construction of knowledge (Fischer, Bruhn, Gräsel, & Mandl, 2002; Hmelo-Silver, 2003).

In this study, students used the NetsBlox (Broll et al., 2017) environment to develop secure robot programs. NetsBlox features a collaborative editor where students can code together in a shared and synchronized workspace while working across multiple computers. Students were co-located and used
their own laptops as they developed their robot programs, dynamically choosing their own roles in accord with findings related to potential issues of forced turn-taking (Tsan, Lynch, & Boyer, 2018).

In order to investigate the collaborative potential of NetsBlox as a flexible, synchronous, co-located platform, we target the following research questions: (1) What programming behaviors as observed via differential sequence mining distinguish groups of students that are successful in the assigned tasks versus those that are not? (2) How did the relative presence or absence of collaboration during a specific project affect student understanding as indicated by learning gains?

RQ1 provides insights into the collaborative programming process students used during the intervention. We hypothesize that groups with greater shared understanding of the task implemented patterns demonstrating systematic development of their code. RQ2 studies the impact of collaboration on learning performance on specific tasks. We hypothesize that groups with greater interactive discourse demonstrated improved collaborative processes targeting a shared understanding.

2. Background

Collaboration is “a coordinated, synchronous activity that is a result of a continuous attempt to construct and maintain a shared conception of a problem” (Roschelle & Teasley, 1995, p.70). Successful collaboration requires the co-construction of knowledge (Fischer et al., 2002; Hmelo-Silver, 2003) between group members along with interactions on progress monitoring, providing constructive feedback, and encouraging the contribution of ideas (Garrison & Akyol, 2013; Grau & Whitebread, 2012). To increase understanding of collaborative problem-solving (CPS) processes, researchers have developed several methodologies, such as after-the-fact reviews (Zimmerman & Pons, 1986) and concurrent think-alouds (Branch, 2013) to gain a deeper understanding of problem-solving processes. However, these methodologies have limitations, such as determining if the student is actually reporting what they thought at the moment and not inferring after the fact, unintentionally biasing the students with leading questions and interrupting the student’s thought processes (Branch, 2013; Zimmerman & Pons, 1986). To overcome such limitations, researchers have utilized students’ naturally occurring collaborative dialogue to provide insight into their collaborative problem-solving processes. Tsan et al. (2018) performed a qualitative analysis of pairs of students working collaboratively, including observations of how students claimed roles and how equally distributed dialogue was among group members. Zakaria et al. (2019) compared collaborating students’ discourse when they either shared a computer or worked side-by-side on separate computers on a shared project, finding no significant differences in the quantity of collaborative talk between the two configurations.

While this dialogue is a rich resource, analysis can be time-intensive and may miss out on relevant information. For example, students may apply signifiers such as “this” or “that” in their discourse, referring to specific program components that are not immediately clear when looking at discourse transcriptions alone. Including students’ action logs can overcome such limitations. Kinnebrew, Segedy, and Biswas (2017) contributed a framework combining analytic measures (model-driven metrics) with patterns derived from students’ action sequences (pattern mining) during model building tasks to understand students’ problem-solving processes and the difficulties they face when working in open-ended learning environments (OELes). Werner, McDowell, and Denner (2013) provided a tool to combine low-level logs into interpretable problem-solving strategies.

However, students typically share one computer during co-located collaborative programming, and this makes it difficult to distinguish one student’s actions from another. A multimodal analysis approach that includes their discourse as well as individual actions in the system allows for a deeper analysis of collaborative problem solving. SLAM-KIT (Noroozi et al., 2019) is recent work aimed at aiding researchers as they combine multiple data streams and seek to better understand collaborative learning. Grover et al. (2016b) present a pilot study for multimodal analysis within a collaborative setting. Vrzakova et al. (2020) considered screen activity, discourse, and body motion as unimodal primitives and extended them to their multimodal combinations.

Multimodal analysis creates an opportunity to further understand collaborative programming. Research on students’ collaborative programming has mostly focused on pair programming. However, some take-aways can apply to other forms of collaborative programming. Shi, Shah, Hedman, and O’Rourke (2019) describe effective collaboration, in the context of programming, occurring when all group members are actively involved in all aspects of the problem process, particularly the construction
of the solution. During successful collaboration, group members share the responsibilities of discussion, planning, and implementation. Additionally, collaborative programming can improve individual programming skills more than solo programming (Braught, Wahls, & Eby, 2011). In this research, we aim to extend our understanding of collaborative programming through the systematic integration of discourse and log analysis to determine how these collaborative factors impact group learning.

3. Methods

3.1 NetsBlox and RoboScape

NetsBlox, an extension of the Snap Programming language (http://snap.berkeley.edu/), is a visual programming environment created to teach students about distributed programming (Broll et al., 2017). NetsBlox allows projects to be shared among group members so that they can synchronously edit and execute various sections of their programs. This allows students to flow between different roles and maintain agency over their own work. Student actions are saved to a database as a particular action type, along with metadata such as the user id, project id, and Unix timestamp. This allows for the compilation of a comprehensive timeline of actions from a student and group perspective.

Communication with a physical robot is enabled through Remote Procedure Calls (RPC). One such RPC, RoboScape (Lédeczi et al., 2019), has a set of specific commands that students can use to communicate with Parallax robots nearly instantaneously. RoboScape, like Zero Robotics (Saenz-Otero et al., 2010) and others, uses a robotics-based framework as an entertaining and educational approach to encourage collaboration through competition. Students work together to strategize for each specific competition and troubleshoot errors in their programs, observing any changes in real-time.

3.2 Study Design and Analysis Framework

Thirty-eight high school students spread across two, week-long summer camps completed our intervention using the NetsBlox and RoboScape environment. Like Karaman et al. (2017), we emphasized robotics software over hardware while incorporating project-based learning and collaborative requirements within a module-based structure. Day one began with an introduction to the NetsBlox environment to establish a baseline of programming proficiency. Day two was the first collaborative day for students. They primarily created simple manual driving programs using RoboScape to increase familiarity with the system and grow comfortable working as a team. Next, they progressively studied more complex cybersecurity techniques, with groups implementing both attack and defense strategies. This culminated in a final competition that required both collaboration and programming skills to succeed while fending off cyber-attacks (Lédeczi et al., 2019; Yett et al., 2020a).

Table 1. Our Analysis Framework

<table>
<thead>
<tr>
<th></th>
<th>Learning</th>
<th>Individual Action Analysis</th>
<th>Sequence Analysis</th>
<th>CPS Discourse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desired Characteristic</td>
<td>Learning gains experienced by all group members</td>
<td>Equal distribution of programming activities</td>
<td>Utilization of strategies identified as supporting program development</td>
<td>Predominantly interactive discourse for a shared understanding of the problem</td>
</tr>
<tr>
<td>Analysis – Entire Intervention</td>
<td>Average normalized change of results in CT and cybersecurity</td>
<td>Actions taken by each group member, Gini coefficient</td>
<td>Differential sequence mining of full sequence of actions</td>
<td>N/A</td>
</tr>
<tr>
<td>Analysis – Tug of War project</td>
<td>Average normalized change of results related to project</td>
<td>Actions taken by each group member, Gini coefficient</td>
<td>Markov chains of actions taken during the project</td>
<td>Application of framework to the discourse of students during the project</td>
</tr>
</tbody>
</table>

The students separated into dyads and triads, choosing their own groups. To be considered a group for this analysis, the students had to work together for at least three of the four collaborative days of the
intervention. Each student also had to complete the pre- and post-tests. Twelve groups were included – six dyads and six triads - spread throughout the two weeks of the camp. The framing of our collaborative analysis approach is presented in Table 1 above.

Data logged during the intervention was queried from the NetsBlox database. We extracted the action data from the database and categorized them as Solution Construction (SC) and Solution Assessment (SA) actions (Kinnebrew et al., 2017). SC actions were subdivided into (1) FillField actions when students changed the values of variables, lists, or robot commands; (2) AddBlock or (3) RemoveBlock actions when students added a new block to their program or deleted an existing block; (4) ConnectBlock or (5) DisconnectBlock actions when students attached a block to existing blocks or removed a block from an existing sequence of blocks, respectively. SA actions were likewise split into (1) StartSimulation actions when students clicked on the “green flag” within the NetsBlox environment to begin running a full program or (2) StartScript actions when students executed only a small portion of their code. Additionally, Idle events occurred when students within a group had a gap of 160 seconds or more between two actions. This value was determined by the 95th percentile of the time intervals.

We applied collaborative discourse analysis (Snyder et al., 2019; Snyder, Hutchins, Biswas, & Grover, 2019) superimposing the ICAP (Interactive-Constructive-Active-Passive) framework of engagement modes (Chi & Wylie, 2014) onto the Weinberger & Fischer (2006) framework component of social modes during knowledge construction. This combined framework allows for analysis of students’ engagement as well as their social modes. The resulting framework consists of eight components - we consider six of them within our analysis. (1) Interactive conflicted-oriented consensus building (ICO) occurs when most group members are participating in the model building and are disagreeing about what needs to be done. (2) Interactive integration-oriented consensus building (IIO) occurs when most group members are participating but no member is taking a strong position on what the next step should be. (3) Interactive quick consensus building (IQ) occurs when most group members are participating in the model building and one member’s suggested contribution is accepted with little to no discussion. (4) Interactive elicitation (IE) occurs when most group members are participating in the model building and one student questions the group. (5) Constructive externalization (CEX) occurs when only one group member is participating in the model building and narrates their actions and thought processes. (6) Constructive elicitation (CEL) occurs when only one student is participating and asks the other non-participating group members a question. Inter-rater reliability was checked by two coders on a 36-utterance segment of discourse and resulted in a Cohen’s kappa value of $k = 0.73$, indicating good agreement.

We answered RQ1 via the “Learning” and “Sequence Analysis” columns of our framework (Table 1). The pre- and post-tests (targeting CT and cybersecurity) were graded using normalized change (Marx & Cummings, 2007) to measure the growth from pre- to post-test for each individual student. These were combined to provide an average normalized change (ANC) based on the number of students in the group, leading to a median split into “Large Improvement” (LI) and “Small or No Improvement” (SNI) groups. We applied Differential Sequence Mining (DSM) to evaluate key strategies implemented by LI and SNI groups (Kinnebrew, Loretz, & Biswas, 2013; Kinnebrew et al., 2015). DSM compares the instance support, or $i$-support (I-$S$), of the compared groups. I-$S$ is defined as the average number of times the pattern was used by each group.

A competition from day three of the intervention called “Tug of War” (ToW) was chosen for further analysis drawing from both the action logs and the discourse of students. In this competition, students fought for control over a robot. Each group would simultaneously send commands to a single robot, attempting to maneuver it towards opposing objectives. This project was chosen because of its highly collaborative nature, as teams of students had to program and strategize together in order to succeed. It was also chosen due to the relation between the sections of code often used by successful teams and two questions on the pre-post-tests related to loops and denial of service attacks. Students learned about and implemented denial of service attacks - and defenses for such attacks - in preparation for the ToW competition, while loops were a necessary component of these attacks.

To answer RQ2 within the ToW context, we required all four columns of the framework presented in Table 1. We selected two groups for analysis – one LI group and one SNI group. Log data was used to compute the Gini coefficient (Dorfman, 1979) as a metric to evaluate individual contributions to group performance. Gini coefficients indicate the share of actions taken by each student, with smaller values corresponding with more equally divided actions. We computed ANC on the specific ToW-related questions as well as determining the action counts (adjusted for number of group members and
the half-day length of the project) and Gini coefficient during this time. We also generated Markov chain (MC) models (Russell & Norvig, 2003) with students’ sequences of log actions to analyze overall group programming sequences. Each state in the state space is represented by an action which occurred at time t, and a transition probability indicates the likelihood for another action to occur at time t+1 with the fitted MC model. Finally, we manually applied our collaborative discourse framework to the discussions of the chosen groups during this project, resulting in counts from each category as well as specifically highlighted discussions related to the programming strategies of each group.

4. Results

Summative results indicated learning gains across CT and cybersecurity (Yett et al., 2020a). Further analysis revealed relationships between average normalized change and results from action log data, the strongest of which corresponded to solution construction actions (Yett et al., 2020b).

4.1 RQ1: What programming behaviors as observed via differential sequence mining (DSM) distinguish groups of students with large improvements during the intervention versus those without?

Using the DSM method, we found action sequence patterns that had statistically significant differences in the instance support (I-S) metric between “Large Improvement” (LI) and “Small or No Improvement” (SNI) groups. Table 2 lists the patterns and the DSM results ranked by the effect size of the differences. The threshold of the s-Support (percentage of groups supporting this pattern) of the DSM was set at the 80% level, and there was no gap allowed between actions.

Table 2. DSM Results

<table>
<thead>
<tr>
<th>Pattern</th>
<th>I-S (LI)</th>
<th>I-S (SNI)</th>
<th>p-value</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ConnectBlock x 9</td>
<td>1.3</td>
<td>4.2</td>
<td>0.02</td>
</tr>
<tr>
<td>2</td>
<td>StartScript, DisconnectBlock, StartScript, RemoveBlock</td>
<td>1.2</td>
<td>0.2</td>
<td>0.02</td>
</tr>
<tr>
<td>3</td>
<td>StartSimulation, DisconnectBlock</td>
<td>6.0</td>
<td>16.7</td>
<td>0.03</td>
</tr>
<tr>
<td>4</td>
<td>StartSimulation x 6</td>
<td>4.0</td>
<td>17.8</td>
<td>0.04</td>
</tr>
<tr>
<td>5</td>
<td>FillField, StartScript, FillField, StartScript, ConnectBlock</td>
<td>1.8</td>
<td>0.7</td>
<td>0.03</td>
</tr>
</tbody>
</table>

SNI groups were more likely to perform a long series of edits to their models. The pattern ConnectBlock × 9 (pattern 1) appeared in 100% of the SNI groups, who also had much larger I-S of this pattern (4.2 instances per group). This depth-first programming approach as opposed to programming in parts has been reported in the literature (Grover, Bienkowski, Niekrasz, & Hauswirth, 2016). Secondly, SNI groups used more solution assessment (SA) actions overall, especially the StartSimulation action once or multiple times in succession (e.g. pattern 4). Though SA actions offer important feedback to construct models effectively, their excessive use implies running simulations without modifications to the models, which indicates an inability to comprehend the information provided by simulations and was linked to SNI students’ overall poorer learning performance.

LI groups showed a more systematic programming approach. Patterns 2 and 5 both could indicate productive troubleshooting behaviors of students testing new blocks or variable values. This is further established by the fact that although SNI groups implemented pattern 3 more (which is similar to the initial sequence of pattern 2), the LI groups may have done so in a more systematic way, as they played the simulation, disconnected a block that may have contained an error, played again to test, and (presumably correct in their hypothesis) removed the block. We hypothesize that the increased use of pattern 3 by SNI groups was due to unsystematic debugging methods used to identify code errors, an indication of ineffective trial-and-error approaches (Murphy et al., 2008). Pattern 5 shows these LI groups changing variable values (FillField) followed by testing those changes depending on the current task at hand (StartScript) and adding the block they were modifying to the full program (ConnectBlock).

4.2 RQ2: How did the relative presence or absence of collaboration during a specific project affect student understanding as indicated by learning gains?
Our case study compares the discourse and log data of an LI group (Group 7) to a SNI group (Group 12) during the Tug of War (ToW) task.

4.2.1 Group 7

Group 7 (G7) was a dyad that worked together from Day 2 through Day 4. They were average in terms of Gini coefficient (0.19) and total actions taken per student per day (364), indicating that they evenly contributed to the projects. They also showed the most improvement among any groups with an average normalized change (ANC) of 0.92. Their results during ToW were similar to their overall scores, with an adjusted 321 actions, Gini coefficient of 0.19, and ANC of 1 (as each student began with missing at least one of the questions on the pre-test but each answered the questions correctly on the post-test).

G7 implemented a few sequences (Figure 1) that relate to our findings from Section 4.1. For instance, transitions such as FillField to StartScript may be associated with the productive troubleshooting behaviors described. They also frequently transitioned between FillField and StartSimulation, which indicated a debugging strategy by tweaking the parameters between tests. This model also indicates potentially new troubleshooting techniques, such as running an individual script for local testing followed by running the entire program. Overall, the group exhibited a variety of behaviors depending on the current needs of the project, with at least one non-recursive transition out of each state. One possible cause for concern is the high likelihood of Idle to StartScript, which could indicate a return to the program after some time spent off-task and an immediate test of a feature that had already been attempted unsuccessfully. However, this may also relate to discussion time and the running of the simulation for further verification. Additionally, students in this group had a 38% likelihood of transitioning from AddBlock to RemoveBlock. Though this is not necessarily indicative of mistakenly adding a block only to remove it, it could signify program changes without proper testing.

Figure 1. Markov Chain Results from Group 7

G7 demonstrated predominantly interactive discourse during programming strategy time segments (Table 3). Coded collaborative discourse indicated that only 32% of utterances coded were constructive, with the group’s discourse targeting mainly interactive integration-oriented consensus building (IIO) and interactive elicitation (IE) in developing their models. In the provided example, group members used explanatory words such as “because” that indicate the group’s ability to reason about their programming actions. This coincides with the literature in that group reasoning and explanatory process are supportive of developing a shared understanding (Fischer et al., 2002; Hmelo-Silver, 2003). Both students logically contributed approximately equally to all Interactive discussions, as interaction is required to achieve such a coding. However, G7-S2 otherwise dominated the conversation, spending more time in a Constructive state attempting to engage their partner while G7-S1 responded tersely or
not at all. G7-S2 also commonly explained their programming and competition strategies to G7-S1 and several outsiders to the group, resulting in high quantities of constructive externalization (CEX).

Table 4. Excerpt from Group 7 Highlighting their use of Programming Strategy

<table>
<thead>
<tr>
<th>Student</th>
<th>Discourse</th>
<th>CPS Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>&quot;my sent&quot; isn't working</td>
<td>ICO</td>
</tr>
<tr>
<td>S1</td>
<td>what did you do?</td>
<td>ICO</td>
</tr>
<tr>
<td>S2</td>
<td>me?</td>
<td>ICO</td>
</tr>
<tr>
<td>S1</td>
<td>why are all these right there</td>
<td>ICO</td>
</tr>
<tr>
<td>S2</td>
<td>because they're for the sent commands</td>
<td>ICO</td>
</tr>
<tr>
<td>S1</td>
<td>why are they under forwards (custom blocks)</td>
<td>ICO</td>
</tr>
<tr>
<td>S2</td>
<td>they're under there so that when we send a command it changes &quot;my sent&quot;. I put them there for a reason</td>
<td>ICO</td>
</tr>
<tr>
<td>S1</td>
<td>well &quot;my sent&quot; is going but the &quot;sent&quot; overall isn't changing</td>
<td>ICO</td>
</tr>
<tr>
<td>S2</td>
<td>no it does see, &quot;change sent by 1&quot; right there</td>
<td>ICO</td>
</tr>
<tr>
<td>S1</td>
<td>no, ok. that's not changing. neither is &quot;my sent&quot;</td>
<td>ICO</td>
</tr>
<tr>
<td>S2</td>
<td>where is our robot? there it is</td>
<td>IIO</td>
</tr>
<tr>
<td>S1</td>
<td>&quot;sent&quot; itself is not changing. right there. the overall &quot;sent&quot;</td>
<td>IIO</td>
</tr>
<tr>
<td>S2</td>
<td>mine is</td>
<td>IIO</td>
</tr>
<tr>
<td>S1</td>
<td>but &quot;my sent&quot; is changing</td>
<td>IIO</td>
</tr>
</tbody>
</table>

4.2.2 Group 12

Group 12 (G12) was a triad that worked together from Day 2 through Day 5. They were average in terms of Gini coefficient (0.20) but took the fewest actions (229) and were the only group to regress from pre- to post-test (ANC = -0.08). At an individual level, two students recorded ANC’s of 0, while one recorded a -0.33 – an additional positive consideration is that the regressing student started with a perfect score and another student only missed one question on each of the pre- and post-tests. Though they took a similar number of actions during ToW after accounting for the length of the session (209), their actions were less evenly divided than usual (Gini = 0.32). Despite this, they saw some improvement in the ToW-related pre-post questions (ANC = 0.67), all from the growth of one student as the other two had answered both questions correctly on the pre-test.

Figure 2. Markov Chain Results from Group 12

An immediate cause for concern upon examination of Figure 2 is the high likelihood of repeated StartSimulation (89%) and StartScript (70%) actions. This coincides with our finding in Section 4.1 demonstrating that SNI groups were likely to implement long sequences of SA actions. Some of this can be explained by the nature of the task, as students often needed to execute at least a few scripts (or the full simulation) in sequence in order to perform well. However, usage at this magnitude indicates repeated examination of the impact of one or a few changes without any modifications in between. This group displayed some positive attributes as well. For example, they rarely transitioned from AddBlock...
to RemoveBlock or RemoveBlock to RemoveBlock, instead showing the desired behavior of StartScript following these changes to their program. On the other hand, the model edit actions and the simulation actions were disjoint: unlike that of Group 7, the MC model of this group had no connection between StartSimulation to any model edit action, and the ConnectBlock to StartScript link that is an important part of a debugging strategy was missing. As a result of this disjoint modeling and testing behavior, this group of students did not seem to be able to build and modify the ToW model successfully.

G12’s discourse (Table 4) is defined by significant use of constructive discourse with little to no explanations for programming changes. For instance, S1 says “as long as you use that to move then we shouldn't get locked out,” with no reasoning. Overall, 45% of the group’s coded utterances were constructive, compared to G7’s 32%. Of the interactive discussions, the group implemented mainly IE, in which one student questions the group. G12-S1 exhibited similar tendencies to G7-S2, taking on a leadership role and generating more discussion than their peers. Likewise, G12-S2 was similar to G7-S1, still contributing to the programming process but not as frequently. One group member, S3, rarely participated in discussions (accounting for less than 10% of the group’s coded utterances). Overall Gini data for the group does indicate that all students normally participated somewhat evenly, so this could just be an outlier caused by external factors upon G12-S3.

Table 4. Excerpt from Group 12 Highlighting their use of Programming Strategy

<table>
<thead>
<tr>
<th>Student</th>
<th>Discourse</th>
<th>CPS Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>as long as you use that to move then we shouldn't get locked out</td>
<td>ICO</td>
</tr>
<tr>
<td>S2</td>
<td>It’s not working</td>
<td>ICO</td>
</tr>
<tr>
<td>S1</td>
<td>Okay [name]</td>
<td>CEX</td>
</tr>
<tr>
<td>S1</td>
<td>so</td>
<td>CEX</td>
</tr>
<tr>
<td>S1</td>
<td>I set this to number 2</td>
<td>CEX</td>
</tr>
<tr>
<td>S2</td>
<td>okay</td>
<td>CEX</td>
</tr>
<tr>
<td>S1</td>
<td>So when I press 2</td>
<td>CEX</td>
</tr>
<tr>
<td>S2</td>
<td>Press 2</td>
<td>CEX</td>
</tr>
<tr>
<td>S1</td>
<td>Actually wait we don’t need number 2 that won’t work. We need the flag.</td>
<td>CEX</td>
</tr>
<tr>
<td>S2</td>
<td>The flag?</td>
<td>CEL</td>
</tr>
<tr>
<td>S1</td>
<td>And it needs to be a forever loop</td>
<td>CEL</td>
</tr>
</tbody>
</table>

5. Discussion and Conclusion

At an overall level, a few characteristics stand out as being common between the two groups. They spent an approximately equivalent amount of their conversation time on interactive conflicted-oriented consensus building (ICO) (26 for G7 and 23 for G12). Though this behavior is argumentative in nature, at these low levels it can often be beneficial in terms of an improved shared understanding and moving the project forward (Weinberger & Fischer, 2006). The two groups also displayed a similar quantity of each kind of constructive discourse and an almost complete lack of interactive quick consensus building. On the other hand, G7 conducted almost twice as many IE exchanges than their counterparts in G12 (110 to 66) along with almost four times as many cases of IIO (81 to 22). Overall, G7 displayed almost entirely positive features from the framework established in Table 1, learning together and showing their collaborative behaviors through Individual Action Analysis, Sequence Analysis, and Collaborative Problem-Solving (CPS) Discourse Analysis. On the other hand, despite some promising trends in Sequence Analysis and Learning during Tug of War (ToW), G12 appeared to struggle to work together and reach the desired level of achievement.

The combination of NetsBlox and RoboScape provides a convenient venue for studying collaborative programming. We contributed new findings in the analysis of programming actions taken by students while collaborating, considering both the distribution of actions amongst group members and the full combined sequence of actions of the group. We also used the available data to highlight two groups as they interacted during a collaborative task. Each of these groups showed Interactive behaviors and students filling specific collaborative roles even without the assignment of predefined roles. The dyad was more successful in terms of sharing actions and showing improvements from pre-post, while the triad seemed to leave one student behind and struggle more. Taken together, these groups
indicate that collaboration on the ToW project could have some impact on pre-post results. However, a more complete dataset would be necessary to validate this claim after addressing a few shortcomings.

One limitation of the system is that network issues can lead to interruptions in collaboration as projects become desynchronized; an integrity check to ensure consistency between the distributed applications could help mitigate this issue. We also allowed groups to choose their own members, possibly skewing results as friends teamed up and worked well together while students that did not know each other had to build their relationship. To somewhat mitigate this disadvantage, lectures on collaboration (Karaman et al., 2017) would be a welcome addition. A final area for improvement is in collecting audio/visual data throughout the intervention, as problems such as microphones being too far away from students and students manually canceling recordings caused us to lose some or all data for certain groups. The combination of these and other factors forced the decision to compare a dyad and triad for our case studies. Ideally, we would compare only dyads or only triads for consistency. We plan to remedy the discussed limitations and introduce additional analysis methods in order to progress the understanding of the effectiveness of our platform.

Acknowledgements

This material is based in part upon work supported by National Security Agency Science of Security Labelt H98230-18-D-0010 and National Science Foundation grants CNS-1644848, CNS-1521617, and DRL-1640199. Any opinions, findings, and conclusions expressed in this material are those of the author(s) and do not necessarily reflect the views of the US Government.

References


Computational thinking and collaborative storytelling in learning Python

Chien-Sing LEEa, Nandakumara VEGANb, Madhavan BALAN NAIRb, James Ean-Huat OOIb
aSchool of Science and Technology, Sunway University, Malaysia
bLee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Malaysia
*chiensingl@sunway.edu.my, madhavan@utar.edu.my

Abstract: Creative and computational thinking are different but complementary. Combining them would bring significant benefits for young people. With interdisciplinarity in mind, FunPlay Code aims to bridge STEM disciplines with the humanities, and the Arts by combining collaborative storytelling and programming. Modelled after Scratch, FunPlay Code encourages users to express experiences in Python codes. These computational concepts/perspectives can be shared, commented, liked, modded collaboratively in a story format. A search function to enable filtering of stories further caters to users’ interests. Functions are developed based on Feature Driven Development methodology. We investigate whether FunPlay Code would be perceived to be easy to use, useful and the likelihood of technology acceptance. User acceptance testing is done remotely with five participants due to the country’s covid-19 Movement Control Order (MCO)/lockdown. Findings are relatively positive. The highest mean score is for social interaction/collaborative storytelling, possibly because the story is fun/surprising yet academic.

Keywords: design thinking, computational thinking, collaborative storytelling, Python, modding

1. Introduction

Universal efforts are put in place to develop policies for secondary and tertiary level education and to develop emerging research priorities in STEM disciplines. Researchers such as Freeman, Marginson, and Tytler (2019), however, note that many young individuals lack the motivation to carry out scientific inquiry during learning as it requires the use of models and processes which to some, can be rigorous and tedious. Thus, this problem needs to be addressed urgently and systematically to foster creative individuals who can create new and inventive technologies for diverse markets and conditions.

Dym, Agogino, Eris, Frey and Leifer’s (2005) propositions for interdisciplinary Engineering Education, focuses on design thinking-based hypothesis formulation/theorizing, towards more viable and systemic human-centered problem-solving and innovation. Such technological narrative should propel society forward in a human-centric impactful manner, compared to a pure programming approach. In parallel with design thinking, computational thinking (Wing, 2006; Brennan & Resnick, 2012) benefits the Arts and humanities through decomposition, pattern recognition, abstraction, and algorithmic thinking. With more practice, computational perspectives and concepts can be reified.

1.1 Problem statement

Schools in Malaysia, a middle-income emerging economy, generally, do not teach programming. Programming is taught only in high school, and even then, often for those in the Science stream. In recent years, however, some children interested in programming would take part in programming classes either informally (online or vendor-based) or formally, such as those organized by the Malaysian Digital Economy Centre (MDeC). Hence, there is a gap between practices in developed countries such as Europe and the US and many emerging economy countries, such as Malaysia.

There are generally two main approaches to computational thinking education. One is non-native (indirectly linked computational thinking-STEM education) whereas the other is native (directly linked). Two exemplary non-native studies to synergize computational thinking-STEM education are generic educational modelling language, e.g. the IMS Learning Design (LD) and Miao, Sodhi, Brouns, Sloep and Koper’s (2008) domain-specific modelling (DSM) approach. DSM synergizes pedagogic experts and technical experts. Pedagogic experts develop notations directly
relevant to the concepts and rules for each domain. In contrast, technical experts create transformation algorithms, which map the models represented in the pedagogy-specific modelling language into computable LD to create context- and needs-specific authoring tools.

Hoppe and Werneburg (2019) also regard data and process structure abstraction as core to “representational flexibility.” However, they mainly focus on mapping simulated models with actual natural phenomena. They utilize “reactive rule-based programming”-type of program analysis to enable learners to begin with situational specifications of action and then further expand these (functional/non-functional requirements) into more standard block-based iterative programs. Their functional/component-based approach allows easier transition between different computational approaches. Consequently, their next study would be towards meta-level programming to complement metacognitive strategies.

More native to computational thinking-STEM education, is Hasan and Biswas’ (2017) domain specific modelling language design with heavier foci on pedagogy while enhancing reusability, interoperability, and rich personalized units of learning.

1.2 Project Objectives

Computational thinking draws on ideas fundamental to computing to solve problems in other fields or domains. The essentials in computational thinking are defining and working with multiple layers of abstractions while understanding their relationships. Wing (2006) focuses on decomposition, pattern recognition, abstraction and algorithmic thinking. Brennan and Resnick (2012) however, focus on much younger learners and thus focus on developing computational perspectives, computational concepts and computational practice.

This paper continues from Lee and Wong (2015; 2018), Lee and Jiang (2019) and Lee and Ooi (2019). Lee and Wong (2015) summarizes design thinking-based research in the creative industries and the development of theorizing and dispositions while Lee and Wong (2018) summarizes research in computing and information systems. In Lee and Jiang (2019), findings indicate that the difference between novices and experts are in perspectives and abstraction mediated by multiple media. In Lee and Ooi’s (2019) conceptual paper, codes are used to express experiences and these codes are in small blocks, emulating functions. Findings indicate the importance of human factors especially familiarity to the task and the need for more examples in the database. The ensuing prototypes aim at developing and reifying computational perspectives and computational concepts through computational practice.

The Web-based version is reported in the 2020 Computational Thinking Conference (Phuan, Lee & Ooi, 2020). Since it is Web-based, an additional future interest is the analogical portability of computational concepts to Internet of Things (IoT) and blockchain. This paper presents the Android-based version of FunPlay Code. The programming concepts/skills we hope to develop is through the language Python. This is due to popular usage among universities in Malaysia. The objective is to test user acceptance of FunPlayCode’s concepts and prototypes. Hypotheses are:

Hypothesis 1: Perceived ease of use of FunPlay Code will positively affect perceived usefulness.

Hypothesis 2: Perceived usefulness of FunPlay Code will positively affect attitude towards using it.

Hypothesis 3: Collaborative storytelling will positively affect user’s attitude towards FunPlay Code.

2. Related work

2.1 Importance of developing interest and habit

Simulated models are often used to externalize and reify human ideas. The Learning Sciences, Educational Technology, Artificial Intelligence communities have progressed far with diverse initiatives aimed at inclusiveness and the Learning Sciences at developing interest and habit. Examples are Lee and Wong’s (2015; 2018) design thinking-based research in Malaysia and Wong, Chan, Chen, Looi, Chen, Liao, King and Wong’s (2020) interest-driven creator (IDC) theory. The latter aims to co-construct a holistic developmental/design framework to guide students to develop their own their learning interests, capability to create, and learning habits. This corresponds with the interest loop,
creation loop and habit loop. The “interest loop” involves triggering interest, immersing interest, and extending interest. Consistent with creativity theories, foci are on developing curiosity, flow, and meaningfulness. The “creation loop” is based on constructionist philosophies, a key tenet in the Learning Sciences. Consequently, interest drives practice and meaningful habits; the “habit” loop.

2.2 Collaborative storytelling

Despite their age or upbringing, everyone undoubtedly has a story to tell. It has helped in communication of sporadic thoughts, ideas and experiences in a structured manner, which ultimately leads to the forming of societies capable of creating stories that span generations. To illustrate the effects of collaborative storytelling, Djerassi’s (1998) experiment with Renga, a collaborative style of Japanese linked-verse poetry genre, comprises a minimum of two stanzas that alternate between different authors, who do not know each other’s identity. Each student contributes a paragraph to his ‘science-in-fiction’ experiment aimed at forming a short story addressing a scientific ethical conundrum. By masking science as fiction, collaborative storytelling transcends conventional learning tools. Renga’s conversational alternating technique, is analogical to extreme programming.

Similarly, to Cao, Klamma and Martini (2008) facts connected with positive user experience can make people retain information easier in a knowledge sharing environment. To bring together different types of thinking, their Personalized Storytelling Environment (PESE) Community of Practice (CoP) allows a user to a) start a story project; b) request other PESE users to join the project; c) search for a story by title or tags; d) search for a story with an algorithm that searches through each user’s profile; e) seek an expert’s help. To motivate mentoring, a user can be promoted to an expert based on number of useful advice. We find these initiatives interesting and inspiring.

Finnish schools (Tenhunen, 2018), Hahai (Tasso, Gervasi, Locchi, A., Sabbatini (2019) and many countries, motivate students towards problem-solving and storytelling in younger grades, the former as young as grade 1. At such a young age, when imagination is unconstrained, fun and learning-by-exploring, creating and collaborating take on new dimensions.

3. System Design

FunPlay Code (Lee & Jiang, 2019; Lee & Ooi, 2019) attempts to provide a platform for young people to explore complex STEM topics in a creative and fun way through collaborative storytelling and coding. This is reflected by the contributor and creator roles in FunPlay Code. The social element engages and stimulates them to have fun and to collaborate and learn with their peers. Requirements are to:

- design and develop an application that can interpret Python code typed using a smartphone interface and display the output.
- enable multiple users to collaboratively contribute to a single story;
- enable story creators to restrict chapter posting access to select users and him/herself only by being able to manually add contributors (initial design);
- develop a filtering function to search stories by typing keywords that match any title or description;
- design and implement a code editing space with syntax highlighter.

4. Software development methodology

To satisfy the fluctuating change in requirements at different phases of development, flexible methods would be more suitable. The agile methodology for this project is the Feature Driven Development (FDD). Development is incremental and revolves around five processes (Figure 1). Hence, it is suitable for a learning environment which emulates Resnick’s (2002) “playground” spirit. Development tools for the project are PyCharm, WebStorm, MySQL 8.0 Command Line Client, Python, Flask, Javascript and React Native.

Figure 1. FDD’s five processes
4.1 System development

Sample screenshots of the developed prototype are presented in Figures 2a, b.

![Figure 2a. Home screen](image)

![Figure 2b. Chapters and floating action button](image)

5. System testing

This twofold testing takes the form of a software engineering testing and a user acceptance testing. The latter is based on the Technology Acceptance Model (TAM). The story that would be used in the test, to comply with the objective of the project, has elements of Mathematics, namely the Fibonacci sequence and Decimal to Binary conversion. It even contains some relatively complex Python functions to test the robustness of the application as partly a Python interpreter. The print function is used generously to output the results of the computed functions and to display the text that were used to describe events of the story and for dialogue. The story draws inspiration from the successful book that is later adapted into a movie, *Hitchhiker’s Guide to the Galaxy* by Douglas Adams.

5.1 Participants

Five consenting participants (3 male and 2 female) with ages ranging from 23 to 25 have taken part in this user testing. The small sample size is because the user testing is carried out during the country’s covid-19 Movement Control Order/lockdown. These individuals have backgrounds in engineering, physiotherapy, programming and mathematics respectively. Due to the small sample size, our study is a case study, similar to a pilot test.

5.2 User testing methodology

Before the date of the User Acceptance Testing, the server of the application is successfully deployed to Google Cloud, as well as the database which was imported into Cloud SQL to enable transmission of data to all the users concurrently. Since the software has to be tested remotely due to the country’s covid-19 Movement Control Order (MCO)/lockdown, a communication line is established with the participants via a WhatsApp group chat.

An APK file containing the actual application with established connection to the server is sent to all the users, ahead of time for the participants to familiarize with the application. They are also politely requested to visit the Info screen for a short tutorial on how to use the app. A copy is also downloaded and installed into the android smartphone of the author to monitor the participants’ actions.

The User Acceptance Testing officially begins when the author is notified by each participant of their online presence on the group chat. All the participants are then asked to login with their unique credentials that are previously assigned. One participant is selected at random to be the creator. The rest play their roles as contributors. The designated creator is tasked to create a story and add the other participants as contributors by identifying their usernames from the multi-select dropdown that would appear once the correct sequence of actions is taken by the creator, in a round-robin fashion. They
would be given access to the script written on Notion, a collaborative multi-purpose note-taking tool, to retrieve the code for each chapter. Afterwards, the other participants are instructed to pull-to-refresh their Home screens and locate the new story that had been created by identifying the title. They then navigate to the Chapters Screen. The entirety of the testing occurred in one session encompassing all the participants as this application is intended to be used collaboratively.

5.3 Technology Acceptance Model instrument/questionnaire

In this project, a questionnaire is utilized to investigate the acceptance of FunPlay Code. The survey items are based on a slightly modified version of the original Technology Acceptance Model (TAM; Davis, 1989) which proposes certain constructs or factors that determine the extent of users’ acceptance. This questionnaire includes 4 items for perceived usefulness, 4 items for perceived ease of use, 5 items for attitude toward using, 4 items for intention to use and an additional 3 items for social interaction to validate the social component of this application. The questionnaire is measured in a 5-point Likert scale that ranges from “strongly disagree” to “strongly agree”.

6. User testing results and analysis

The user testing results and analyses are presented in Table 1.

Table 1. User testing results and analyses

<table>
<thead>
<tr>
<th>Construct</th>
<th>Questionnaire Item</th>
<th>Mean</th>
<th>Cronbach’s α (Full scale)</th>
<th>Cronbach’s α (Subscale)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived Usefulness</td>
<td>PU1</td>
<td>3.4</td>
<td>0.0135</td>
<td>0.29</td>
</tr>
<tr>
<td>PU2</td>
<td>3.6</td>
<td>0.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PU3</td>
<td>3.0</td>
<td>0.0134</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PU4</td>
<td>3.4</td>
<td>0.0204</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>3.28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived Ease of Use (PEU)</td>
<td>PEU1</td>
<td>2.8</td>
<td>0.7427</td>
<td></td>
</tr>
<tr>
<td>PEU2</td>
<td>2.8</td>
<td>0.7705</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEU3</td>
<td>2.8</td>
<td>0.6935</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEU4</td>
<td>2.8</td>
<td>0.6821</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>2.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attitude Toward Using (ATUS)</td>
<td>ATUS</td>
<td>3.4</td>
<td>0.8382</td>
<td></td>
</tr>
<tr>
<td>ATUS</td>
<td>3.4</td>
<td>0.7049</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATUS</td>
<td>3.4</td>
<td>0.7049</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATUS</td>
<td>3.4</td>
<td>0.7049</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>3.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intention of Use (IU)</td>
<td>IU1</td>
<td>4.2</td>
<td>0.8374</td>
<td></td>
</tr>
<tr>
<td>IU2</td>
<td>4.2</td>
<td>0.7177</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IU3</td>
<td>4.2</td>
<td>0.7177</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IU4</td>
<td>4.2</td>
<td>0.642</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>4.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Interaction (SI)</td>
<td>SI1</td>
<td>2.4</td>
<td>0.5235</td>
<td></td>
</tr>
<tr>
<td>SI2</td>
<td>2.4</td>
<td>0.4257</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI3</td>
<td>2.4</td>
<td>0.4257</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>2.4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- For all of the constructs, the total average score is over 3.45 which indicates that the users perceive the application quite favourably.
- Social Interaction at 4.13 is the construct with the most positive response. These responses imply that the users are attracted to the storytelling and social aspect the most.
- Among the subitems, the best response with a mean score of 4.4 is for ATUS5, i.e. the fifth item in Attitude towards using FunPlay Code, “I like the storytelling aspect of FunPlay Code.”
- Cronbach alpha values are positive but sample size is too small.

7. Conclusion

In conclusion, the FunPlayCode collaborative storytelling mobile application using Python code has
potential. However, collaborative storytelling in a non-academic manner is attractive to young coders, if the story is fun/surprising. Despite the remote testing, the sharing of screenshots and the Info screen are helpful as guides whenever a participant faces any complications. There are however, limitations to the study e.g. small sample size. Moreover, the learning curve would reduce if there are more examples to build on. Future work may include linking to social media accounts to invite friends to code with.

Acknowledgements

We thank Dr. K. Daniel Wong and Universiti Tunku Abdul Rahman for introducing/supporting design-computational thinking since 2013 when the first author was a Faculty at UTAR, and the anonymous reviewers’ detailed feedback. This paper is extended from Nanda’s final year project thesis.

References

Investigate the Influence of Interactive Immediacy on Collaborative Knowledge Construction in Online Discussions

Wanyu ZHANG*, Luyue LI†, Lei ZHANG*, Fan YE* & Ning MA*

*School of Education Technology, Faculty of Education, Beijing Normal University, China
†horseming@bnu.edu.cn

Abstract: Online learning platforms can be divided into two categories according to the interactive immediacy, platforms with weak immediacy and platforms with strong immediacy. This study selected one of these two kinds of platforms to investigate the differences of interaction characteristics and collaborative knowledge construction in online discussions. A total of 46 first-year master students’ online discussion texts were collected and coded as Pena-Shaff’s and Gunawardena's coding scheme, which led to the following conclusions: (1) online discussions with strong immediacy facilitated the formation of interactions with multiple rounds and periods of participation; and (2) online discussions with strong immediacy had more categories of interactions and deeper levels of collaborative knowledge construction than those with weak immediacy. In addition, this study gave some recommendations about online discussions in both platforms to facilitate collaborative knowledge construction.

Keywords: Collaborative knowledge construction, interactive immediacy, online discussion

1. Introduction

An important factor influencing online discussions is the immediacy of online learning platforms, according to which this study classifies platforms into two categories. The first is the web-based platforms with weak immediacy, such as Wikipedia, Moodle, Learning Cell System, etc. This kind of platform with different modules not only provides services to students but integrate a variety of teaching functions, such as supporting students' collaborative learning (Hu Li et al., 2016). The second category is the social media applications for mobile with strong immediacy such as Twitter, Facebook, WeChat, etc., they are convenient and extremely popular among college students (Demir, 2018; Xue & Churchill, 2019). When applied to teaching, it can closely connect learners and teachers and support the interaction between them and promote students’ social learning (Muls et al., 2020).

It can be seen that online learning platforms with different interactive immediacy may provide different support to students, and its impacts on online discussions, such as the online interaction and the level of collaborative knowledge construction remain to be investigated.

2. Literature Review

There are numbers of studies that have discussed online discussions with varying degrees of interactive immediacy. Students interact more frequently in online discussions when the interactive immediacy is strong, especially social and emotional interactions. Frequent interactions can also increase student motivation and responsibility (It-analysis, 2001); when the interactive immediacy is weak, students can express more complex, reflective views in online discussions (Hrastinski, 2008). However, this does not mean that strong immediacy has no effect on the cognitive development. Researches have shown that in synchronous learning environments with strong interactive immediacy, timely feedback can immediately correct students, consolidate learned knowledge, thus facilitate group decision-making, brainstorming and analysis (Chen et al., 2005); also, students’ collaborative knowledge construction can be facilitated through real-time self-regulation and collaborative regulation (Lee et al., 2017).
However, it is not sure whether different interactive immediacy can influence the degree of cognitive development of online discussions.

Researchers usually analyze the degree of cognitive development in collaborative learning by assessing students' collaborative knowledge construction (Puntambekar, 2006), and online discussion designed for this study is also online collaboration. The current researches have not discussed the collaborative knowledge construction of students when interactive immediacy is different, so this study chose two online discussion platforms with different interactive immediacy, WeChat app with stronger immediacy and Learning Cell System with weaker immediacy to investigate two research questions:

Q1: What’s the difference of interactive characteristics when interactive immediacy in online discussions is different?
Q2: Is there difference in levels of collaborative knowledge construction when interactive immediacy in online discussion is different?

3. Methodology

3.1 Participants

46 first-year master students participated in this experiment, whose majors are education-related: educational technology, curriculum and pedagogy and educational principle. They have similar professional backgrounds and are proficient in Learning Cell System and WeChat.

3.2 Procedure

Firstly, all students took part in the same course "Technological foundations of education". After learning the theme of "STEM Education Cases, Connotations and Teacher Preparation", they were given a related discussion question “What is the dilemma of STEM education development in China in your opinion?”

Students were randomly divided into two groups SI (strong immediacy) and WI (weak immediacy), with each group 23 students. WI discussed this question on Learning Cell System and SI discussed on WeChat. After a week, the discussion messages were collected and coded by four researchers. Two researchers analyzed the messages of SI, and the others analyzed the messages of WI.

3.3 Instrument

Pena-Shaff's and Gunawardenaet's coding schemes are widely used to analyze interaction features and collaborative knowledge construction in collaborative learning, so this study used their coding schemes to analyze discussion messages. Pena-Shaff and Nicholls (2004) divided discussion into interactive messages and non-interactive messages, including a total of 11 categories of messages: Question, Reply, Clarification, Interpretation, Conflict, Assertion, Consensus Building, Judgment, Reflection, Support and Other. Gunawardena divided the process of knowledge construction into five phases, including Sharing/Comparing, Dissonance, Negotiation/Co-construction, Testing Tentative Constructions, and Statement/Application of Newly-Constructed Knowledge (Gunawardena, Lowe, & Anderson, 1997). This study used the former to analyze the interactive characteristics and the latter to analyze the layers of knowledge construction in two groups. Researchers coded the messages as the definitions of these categories and phases. When coding as Pena-Shaff's coding scheme, for instance, the message of "......The introduction of STEM is a good way to facility innovation, but there are several obstacles: Firstly, it’s the gap in professional development of teachers. STEM teacher standards have not been established and STEM teacher training programs are lack......" was coded as “Clarification”, "......We can leverage the market, advising researchers to work with companies to develop adaptable STEM products and resources to improve STEM courses......"was coded as “Interpretation” and "agree with you! " was coded as “Support”. After negotiation and calculation, Pena-Shaff's and Gunawardenaet's coding consistency were respectively 0.87 and 0.85.
4. Results and Analyses

4.1 Number and Length of Messages

The numbers of discussion messages in two groups are shown in Table 1. There were 65 messages in WI and 36 messages in SI. Average message length in SI was longer than WI. Possible reasons for this phenomenon were: (1) the number of messages which students replied in WI accounted for two-thirds and they were generally shorter; (2) students in SI focused more on describing their own views, resulting in an increase in message length.

Table 1
The Number of Messages in Group WI and SI

<table>
<thead>
<tr>
<th>Group</th>
<th>Message Number</th>
<th>Total Length</th>
<th>Average Message Length</th>
<th>Longest Message Length</th>
<th>Shortest Message Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>WI</td>
<td>65</td>
<td>13443</td>
<td>206.8</td>
<td>879</td>
<td>22</td>
</tr>
<tr>
<td>SI</td>
<td>36</td>
<td>8047</td>
<td>223.5</td>
<td>700</td>
<td>5</td>
</tr>
</tbody>
</table>

4.2 Primary analyses

4.2.1 Analyzing Interaction Characteristics

As for the first question, this study analyzed interactions during discussion using message maps and Pena-Shaff’s coding scheme.

The message maps can be seen in Figure 1 and Figure 2. The first line of the message map is the time stamp, DT represents the discussion topics, such as figure "(1)" is discussion ID. The green ellipse indicates the initiation message, and the white is the follow-up message, the figure in ellipse is student’s ID, the line between two ellipses represents discussion thread of the two students, and the arrow points to the respondent. As shown in Figure 1, student #2 posted a message (2) in WI on 5th Dec, student #3 commented on the post on 6th Dec, and then student #2 responded to this comment on the same day.

According to message maps, it can be seen that:

(1) Discussion threads showed differences in time span. In WI, students usually replied to earlier messages, so these messages received more replies. For example, students #18, #19 and #6 were online on 11th Dec, they would check and respond to the message sent on 5th Dec, which were replied six messages for each and also replied most. Whereas students in SI usually responded to latest messages, as shown in Figure 2. Therefore, there’s a larger time span between discussion threads in WI.

(2) Students in SI have more interaction rounds than WI. Students in WI interacted with each other up to 1 round, such as on 5th and 6th Dec, when student #2 interacted with student #3, while the interaction rounds were up to 3 in SI, as shown in Figure 2. At 21:03 - 22:48 on 10th Dec, students #20 and #21 interacted three rounds.

(3) Students in SI were more likely to discuss at different times than in WI. Only four students discussed in different time in WI, and they were students #1, #2, #3 and #6 whereas there were six students’ discussing in different time in SI, such as students #1, #3, and #20.

Regarding as interaction categories coded as Pena-Shaff’s scheme, you can see relevant data in Table 2, this study compared the percentage of each category in two groups, findings are as follows:

(1) The most interaction categories were same in two groups, they were “Clarification”, “Interpretation” and “Support”. And the number of “Clarification” was largest in two groups, with respectively 46.32% in SI and 35.94% in WI and the percent of SI was 10 more than WI. The percent of “Interpretation” and “Support” of WI exceeded SI respectively 5% and 10%.

(2) There’s 11 interaction categories in SI but only 7 in WI, and the percent of “Question”, “Conflict”, “Assertion”, “Consensus Building”, and “Reflection” was actually low even in SI. “Question” shares respectively 2.11% and 2.34% in group SI and WI. In SI, “Conflict”, “Assertion” and “Consensus Building” all accounted for 1.05 percent and the percent of “Reflection”
was 2.11%, while there’s no reflective discussion in WI. There were also other categories only can be found in SI sharing relatively high, such as “Response” with 6.32%.

4.2.2 Analyzing Levels of Collaborative Knowledge Construction

As for the second question, this study used Gunawardenaet’s coding scheme to analyze the levels of collaborative knowledge construction in two groups. As shown in Table 3, by comparing the percent of messages at each level of collaborative knowledge construction, it can be found that:

1. Group SI had a deeper level of knowledge construction than WI with reaching the stage PH3. Although the number of messages of SI was less than WI, it reached a deeper level of interaction.

2. Neither WI nor SI reached the deepest stage of collaborative knowledge construction, and the two groups respectively accounted for 69.44% and 95.38% on stage PH1.

5. Conclusions and Future Work.

5.1 Conclusions and Suggestions

After investigating the interactive characteristics and levels of collaborative knowledge construction with different immediacies in online discussions, this study found:

1. Strong immediacy helps form multiple rounds, multi-period participation, and multiple categories of interaction.

   Students in SI can ask questions, respond to others and get feedback timely. Timely information exchange can trigger intense discussions among students, which generates more interaction categories and more interaction rounds. Also, platform with strong immediacy allows students to engage in discussing quickly at different time, whereas students in WI hardly get timely feedback and they are more inclined to describe their own views.

2. Strong immediacy is conductive to a deeper level of collaborative knowledge construction.

   As the first conclusion suggests, timely feedback from strong immediacy provokes thought from students and thus deepens the level of collaborative knowledge construction. But the time interval between posting and replying of weak immediacy is longer, so it is difficult to arrive deep collaborative knowledge construction during the short experiment period.

   In short, platforms with stronger immediacy are conductive to interacting and deepening collaborative knowledge construction in online discussions, although knowledge construction is not so ideal in this experiment. In response to this problem, the roles of teachers and opinion leaders can be
### Table 2

**Categories of Interaction in Two Groups**

<table>
<thead>
<tr>
<th>Group</th>
<th>Question</th>
<th>Response</th>
<th>Clarification</th>
<th>Interpretation</th>
<th>Assertion</th>
<th>Conflict</th>
<th>Consensus Building</th>
<th>Judgment</th>
<th>Reflection</th>
<th>Support</th>
<th>other</th>
</tr>
</thead>
<tbody>
<tr>
<td>WI(%)</td>
<td>3</td>
<td>0</td>
<td>46</td>
<td>31</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>11</td>
<td>0</td>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>SII(%)</td>
<td>2.34</td>
<td>0</td>
<td>35.94</td>
<td>24.22</td>
<td>0</td>
<td>1.56</td>
<td>0</td>
<td>8.59</td>
<td>0</td>
<td>24.22</td>
<td>3.13</td>
</tr>
</tbody>
</table>

### Table 3

**Levels of Collaborative Knowledge Construction In Two Groups**

<table>
<thead>
<tr>
<th>Phase</th>
<th>SII Message Number</th>
<th>SII Percentage</th>
<th>WI Message Number</th>
<th>WI Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>PH1: Participants share information and ideas with each other and describe the topics discussed</td>
<td>25</td>
<td>69.44%</td>
<td>62</td>
<td>95.38%</td>
</tr>
<tr>
<td>PH2: Participants identify and analyze inconsistencies in ideas, concepts or descriptions and deepen their understanding of the problem</td>
<td>6</td>
<td>16.67%</td>
<td>3</td>
<td>4.62%</td>
</tr>
<tr>
<td>PH3: Participants construct knowledge Collaboratively through meaning negotiation</td>
<td>3</td>
<td>8.33%</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PH4: Participants examine and modify newly constructed ideas</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PH5: Participants reach agreement to apply newly constructed knowledge</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
considered. Previous studies show teachers' behavior can affect the quality of online discussions. Scaffolds and novel perspectives from teachers help students think deeply. Guidance and feedback from teachers also enhance their enthusiasm (Yang, Lv, Wang, & Wang, 2009; Yu, Li, & Wang, 2010). Also, opinion leaders often form direct or indirect social relationships through online communication, causing ripple effects to others (Li, Ma, Zhang, Huang, & Kinshuk, 2013). Therefore, deepening the levels of knowledge construction through the intervention of opinion leaders is worth investigating.

It is participants’ course tasks that may result in the inconsistency with the previous research (Hrastinski, 2008). In addition, this experiment shouldn’t be too long because it is only a theme of this course, which helps platforms with strong immediacy perform better.

5.2 Limitations and Future Work.

This study investigated the interactive characteristics and levels of collaborative knowledge construction in online discussion with different immediacy. The small sample and data may cause that our conclusion cannot be open to more conditions. Questionnaires and interviews can be considered to analyze factors that affect collaborative knowledge construction in these two platforms.

It is inferred that types of discussion question can also lead to different results. The question in this study was open and moderate. How sharp and challenging questions work and whether different types of questions will lead to different results can be explored.

Acknowledgements

We would like to thank all the people who prepared and revised previous versions of this document.

References

Supporting Collaborative Inquiry for Geography and Teamwork Learning: The Techno-Pedagogical Design of My Groupwork Buddy for Geography

Elizabeth KOHa*, Yi-Huan TEEa, Dhivya SURESHA, Betsy NGa, Helen HONGab, Tricia SEOWa

*aNational Institute of Education, Nanyang Technological University, Singapore
bMinistry of Education, Singapore

Abstract: In computer-supported collaborative learning, many tools either focus on the cognitive or the social aspect but not both. With the increasing importance of nurturing students holistically encompassing academic and social skills, the later known commonly as 21st Century competencies, systems should therefore provide affordances that cater to these dual aspects. However, it is challenging and there are limited guidelines on designing such integrated tools. To fill this gap, this paper describes the development of a techno-pedagogical system, My Groupwork Buddy for Geography, for collaborative inquiry that intends to deepen students’ knowledge and understanding of geographical topics as well as grow their teamwork. Two key design considerations are synthesized relating to the whole techno-pedagogical approach and the motivational needs of the student. The paper subsequently articulates these considerations in the design and implementation of two trial cycles of My Groupwork Buddy for Geography involving Secondary Three students’ field-based geographical inquiry. Specifically, three techno-pedagogical design principles are elaborated on. In addition to theoretical contributions, this work discusses practical implications and offers guidance towards designing tools that build the dual complementary goals important for learners now and in the future.

Keywords: design, group inquiry, geography, teamwork, motivation

1. Introduction

In Geography, team collaboration is an essential part of the curriculum, as students have to work in teams as part of geographical inquiry. With the ubiquity of technology, students and teachers have adopted an assortment of tools to support them in this collaborative inquiry. However, each tool has specific affordances as well as limitations. In computer-supported collaborative learning, many tools either focus on cognitive awareness or social awareness but not both (Janssen & Bodemer, 2013). It is challenging and there are limited guidelines for tools to support both cognitive and social awareness. In the same vein, many tools for collaborative geographical inquiry emphasize the cognitive aspect, e.g., to build the geographical knowledge of the learner, and do not equally encourage the social aspect, e.g., the teamwork and collaboration of the learner. Moreover, with the current emphasis of future-ready learners who are competent in both academics and soft skills, it is therefore important that collaborative inquiry tools provide opportunities for learners to develop these dual competencies.

This paper describes the development of a techno-pedagogical system, My Groupwork Buddy for Geography that intends to deepen students’ knowledge and understanding of geographical topics as well as growth in the 21st Century competency of teamwork. A design-based research approach is employed over two trials for Secondary Three (15-year-old) students in Singapore. The project also addresses students’ psychological needs, which is often not as much mentioned in the literature in system design. A third contribution is that it theoretically and practically identifies and describes techno-pedagogical design principles for collaborative inquiry that meets the dual complementary goals of academic excellence and teamwork, related to cognitive and social awareness tools respectively.
The paper begins with a literature review of systems that support collaborative inquiry and design considerations. Next, it elaborates on key design principles of the system before closing with the benefits and constraints of the techno-pedagogical system design.

2. Literature Review

2.1 Systems that Support Collaborative Inquiry

Collaborative inquiry is the process of investigating a certain question with a group of people and there are systems that support such inquiry in Geography as well as generic tools customized for group inquiry. For the first category, an example is nQui re (Sharples et al., 2015) designed to support the whole inquiry process, allow primary data collection, and allow teachers to orchestrate inquiries across multiple contexts such as classroom, field, and home (Scanlon, Anastopoulou, Kerawalla, & Mulholland, 2011). While this is a notable system, it did not focus much on the teamwork process.

Generic software has also been used to support collaborative inquiry in Geography such as learning management systems (LMSs). As noted by several studies (e.g., Scanlon et al., 2011), these generic software tools do not support the whole collaborative inquiry process and/or require additional scaffolds to more effectively and conductively support the process for students and teachers.

Lastly, group awareness tools can provide information about group members to support individual and collaborative learning (Janssen & Bodemer, 2013; Schnaubert, Heimbuch, Erkens, & Bodemer, 2019). They can be classified into cognitive group awareness tools (cGATs) or social group awareness tools (sGATs). Most cGATs enable comparison between the learner’s own cognitive information with that of other group members while sGATs allow learners to compare their own behavior to that of others. While both are required for effective coordination, researchers have tended to focus singularly on either one (Janssen & Bodemer, 2013). Similarly, Schnaubert and colleagues (2019) highlight that cGATs are seldom used for other purposes although the versatility of cGATs has potential in guiding collaborative inquiry. They add that there are no existing guidelines or overarching framework to support educators in designing group awareness tools for their teaching needs.

2.2 Techno-Pedagogical Design Considerations

In developing techno-pedagogical tools that support the collaborative inquiry of Geography that specifically foregrounds curriculum content and process as well as collaboration and teamwork aspects, the following design considerations are surfaced.

2.2.1 Tools Need to Foreground and Support the Whole Collaborative Inquiry Process

Scanlon et al. (2011) and Sharples et al. (2015) highlight the importance of systems that encompass the whole inquiry process, rather than piecemeal support of certain aspects only. In foregrounding the whole process, the curricular content as well as the collaborative aspects have to be made explicit. This explicitness helps emphasize to learners what is important of the process of their learning. This also refers to both the cognitive and social group awareness that are important for coordinating collaborative activities effectively (Janssen & Bodemer, 2013).

2.2.2 Tools Must Support Learners’ (Students) Psychological Needs

This principle reiterates the importance of student-centeredness in ensuring that students’ psychological needs are addressed. Drawing from self-determination theory (SDT), students have three basic psychological needs, namely autonomy, competence and relatedness (Deci & Ryan, 1985; Ng, Liu, & Wang, 2015). To facilitate students’ psychological needs, teachers must create a need-supportive environment that fosters autonomous motivation such as identifying and fostering students’ intrinsic motivation by offering options; fostering interest with respect to learning; providing rationale and informational feedback; as well as encouraging self-regulated learning (Reeve & Cheon, 2016).

Extensive studies in the SDT literature have provided the benefits associated with learners’ need satisfaction and teacher’s autonomy support. Although past studies were mainly focused on autonomy-supportive teaching in classrooms or face-to-face instruction, it is possible that technologies can be designed to meet these psychology needs of learners. According to a recent study on an online
pedagogical assessment, the findings suggest students experienced a sense of self-determination when engaging in online quizzes (Chemsi, Sadiq, Radid, & Talbi, 2020).

3. Study Context and My Groupwork Buddy for Geography Techno-Pedagogical Design

This project utilizes the design-based research approach over two trial cycles for Secondary Three students’ field-based geographical inquiry. This form of fieldwork inquiry in teams is commonly known as geographical investigation (GI) in Singapore. A team of researchers, web developers, teachers, and education ministry officers were co-designers in the project and also sought students’ voices throughout the trials. The following describes key techno-pedagogical design principles that were conceptualized and developed over the project, with an emphasis on the finalized design of Trial 2.

3.1 Collaborative Inquiry Model to Support the Whole Process

In view of design consideration #1, to guide the collaborative inquiry in Geography and meet the dual complementary outcomes of deepening knowledge in geographical topics as well teamwork growth, the project designed a model that integrates the curriculum’s geographical inquiry approach with a teamwork pedagogy, the Team and Self Diagnostic Learning (TSDL) pedagogical framework (Figure 1a). In essence, this integrated model combines the unique formative assessment approach of teamwork, which provides on-demand teamwork feedback, with an established geographical pedagogical framework, to encourage dual complementary outcomes. The model was also linearly visualized as a navigation pane on the system for easy referral and clickable guidance to parts of the system with respective online spaces for the activities (Figure 1b).

The Geography curriculum in Singapore employs the geographical inquiry approach as its main pedagogy. Similar to inquiry-based learning, this approach comprises four key aspects, namely question-driven, evidence-based, reflection and knowledge construction. These aspects are represented in a cyclical inquiry process: *sparking curiosity, gathering data, exercising reasoning* and *reflective thinking*. On the other hand, TSDL is a digital formative assessment approach to teamwork (Koh, Hong, & Tan, 2018). Informed by pedagogies such as experiential learning and the learning analytics process model, TSDL relies on four staged mechanisms: 1) *team-based concrete experiences* that involves students engaging in collaborative activities to gain understandings of working with team members; 2) *self and team awareness building* through visualizing self and peer ratings of teamwork behaviors; 3)
team and self reflection and sensemaking where students reflect and set goals based on the insights from the visual analytic; and 4) team and self growth and change for students to monitor teamwork goals set.

The collaborative inquiry model for geography and teamwork learning integrates the four phases of the geographical inquiry process with two TSDL cycles. As seen from the first quadrant in Figure 1a, this is the phase, sparking curiosity, where students have to work in their team to conduct preliminary research as well as construct the guiding question. After this milestone of coming up with the guiding question and hypothesis as a team, students have gained enough concrete experiences (TSDL stage 1) and can proceed to TSDL stage 2, where they will build self and team awareness through self and peer ratings about aspects of their recent teamwork experience. This information is harnessed as a form of feedback for students via a visual representation, for students to reflect on. This provides students with a greater metacognitive awareness of the teamwork aspect, in addition to the taskwork aspect of the geographical inquiry process.

Following this, students will gather data and exercise reasoning (next two phases of geographical inquiry) with specific activities such as representing data. Students are reminded to monitor their teamwork progress through status checks (part of the last stage of TSDL). This makes it more visible for students and provides opportunities for them to regulate their individual and team learning behaviors. Finally, during reflective thinking where students evaluate about their geographical learning, the TSDL cycle can begin again, allowing them to consciously assess their teamwork.

3.2 Student Collaborative Space with Choices and Scaffolds

In developing the tool, through having one electronic platform for GI, students can more easily engage in geographical tasks and team dialogue to construct joint understandings of subject content and teamwork, in line with consideration #1 (Janssen & Bodemer, 2013). Moreover, a collaborative space for students was implemented in Trial 2, via the collaborate function (Figure 2). This allows students to represent their joint understanding of their knowledge visibly in their team through the co-writing space, in addition to the team chat and team reflection space.

Based on Trial 1 feedback from students, more explicit support for student’s psychological needs through the system design was highlighted. Thus, in Trial 2’s redesign, SDT motivating principles were more purposefully incorporated (consideration #2) such as take the students' perspective, offer activities to vitalize psychological needs (competence, autonomy, and relatedness), provide explanatory rationales and communicate using informational language (Reeve & Cheon, 2016).
Collaborative documents for each student team is organized in three sections (Figure 2). Live documents displays all the “active” documents for each team to work on via clicking "Edit". Figure 3 shows a sample working document. It starts off blank, providing flexibility in what students would like to create and edit, and also for teachers to insert guiding questions as scaffolds. On the right of the document, this online presence provides opportunities for students to interact with each other.

Teams had the choice of when to submit the document for review, before its due date. To help maintain learners’ motivation during the GI, instead of one large report that is due at the end, chunking was applied to break down the report into bite-sized deliverables with specific guiding questions that are due at certain timepoints. Also, scaffolds provided in the form of specific guiding questions for each smaller deliverable meant that explanatory rationales were provided at every step of the GI, so that students are clear about their deliverable. This helps to fulfil the motivating principle of offering students a clear standard to work towards, so as to fulfil one of their psychological needs, competence.

3.3 Collaborative Inquiry Content with Learning Prompts

In line with design consideration #2, learning prompts for geographical and teamwork content were added in the forms of hints and tips. Geographical hints were indirect clues on the elements of the GI content (relating to declarative knowledge), to engage students in deeper thinking, whereas tips related to procedural knowledge, and were crafted to relieve students’ pain points at certain parts of the GI. An example tip is provided at GI Stage 2, Construct hypothesis/guiding question, “A little time spent here will save you a lot of time later”. Taking the student’s perspective, this tip highlights the importance of this step by explaining the rationale of time saved. For teamwork learning, teamwork tips were designed to be relatively generic to apply to most teamwork situations and relate to the dimension of teamwork that students wanted to improve on. Similar to the prompts for the geographical content, the tips employed motivating strategies such as taking students’ perspective and using informational language.

4. Discussion and Conclusion

To what extent can tools be designed to support both geographical as well as teamwork learning in collaborative inquiry? This paper has summarized two key design considerations relating to the whole techno-pedagogical approach and the needs of the student. Furthermore, these considerations were
developed in Trial 2 of the project. The collaborative inquiry model that integrated both the geographical inquiry approach as well as TSDL in its very alignment highlights how cognitive and social awareness tools can be combined in a single system, which is a contribution to the field. We are cognizant that the current collaborative inquiry model is depicted as linear in nature. Authentic collaborative inquiry could be more connected, iterative and intertwined (Scanlon et al., 2011). The relatively simplistic illustration was for easier understanding of the collaborative inquiry process, which already integrated the intricacies of two other frameworks. That said, during the implementation, teachers did elaborate to their students that iterative nature of collaborative inquiry.

Second, in designing the collaborative space for students, care was taken to provide flexibility and choice for them, balanced with structure and scaffolds for their guidance. Third, the content in the system relating to the outcomes of geographical knowledge and teamwork learning were enhanced with learning prompts intended to motivate their learning in these respective areas. Through these two designs, the system emphasized and explicitly addressed students’ psychological needs. This incorporation of SDT-based motivational principles remains relatively novel and further work could validate the framework for examining motivational dynamics and experiences of learners.

Through the synthesis and design of My Groupwork Buddy for Geography, this paper has theoretically identified and practically developed three techno-pedagogical design principles for collaborative inquiry. This contributes to the academic literature in learning design and broadens much of extant literature on academic content towards an equal importance on softer skills such as teamwork. Additionally, this study extends the existing collaborative inquiry and SDT literature in the field of educational research. In all, providing guidance towards designing tools that build the dual complementary goals important for learners now and in the future.

**Acknowledgements**

This study was funded by Singapore Ministry of Education (MOE) under the Education Research Funding Programme (DEV 03/17 EK), administered by National Institute of Education (NIE), Nanyang Technological University, Singapore. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Singapore MOE and NIE.

**References**


Improving pre-service teachers’ ICT-integrated lesson design through formative peer feedback

Wenli CHEN*, Zhongling PI†, Jesmine S. H. TAN*  
*National Institute of Education, Nanyang Technological University, Singapore  
†wenli.chen@nie.edu.sg

Abstract: Formative peer feedback plays a critical role in teachers’ collaborative lesson design, which enables teachers to reflect on and continuously revise their lesson design by receiving and providing concrete comments and suggestions on how to improve it. Past research has yielded mixed results on the effect of formative peer feedback on learning, partially because these studies used different types of formative peer feedback with different levels of specificity of the information provided in the feedback. This study examines the relationship between the type of formative peer feedback, the specificity of information provided in the formative peer feedback, and the quality improvement of the ICT-integrated lesson design by a group of pre-service teachers’ in Singapore. The results show that the quality of ICT-integrated lesson design was significantly improved through the formative peer feedback. Furthermore, it is found that positive affective feedback and cognitive questioning feedback with specific explanation facilitated the improvement of the ICT-integrated lesson design whereas cognitive scaffolding feedback with specific explanation hindered the improvement of the ICT-integrated lesson design. The implications on how formative peer feedback affect pre-service teachers’ ICT-integrated lesson design are discussed.

Keywords: Formative peer feedback, ICT integration, TPACK

1. Introduction
With the rapid development and diffusion of technologies, the integration of Information. Communication and Technology (ICT) to facilitate 21st century learning has become pervasive in educational institutes globally (Tondeur, van Braak, Sang, Voogt, Fisser, & Ottenbreit-Leftwich, 2013). The ICT integrated lesson design poses many challenges to teachers. Novice teachers often feel that they are not well-prepared to effectively use ICT for teaching and learning and ICT is under-used in their classrooms (Sang, Valcke, van Braak, & Tondeur, 2010). Previous studies have shown that a crucial factor affecting teachers’ integration of ICT for teaching and learning is their experiences in technology-enhanced learning during their pre-service teacher training programs (Drent & Meelissen, 2008; Tseng, Cheng, & Yeh, 2019).

To prepare pre-service teachers for effective ICT integration, teacher education programs need to help them build Technological Pedagogical Content Knowledge (TPACK) (koehler & Mishra, 2009). TPACK emphasizes the integrated use of technology, pedagogy and content knowledge for effective technology integration (Reyes Jr, Reading, Doyle, & Gregory, 2017). It aims at developing appropriate, context-specific strategies and representations for ICT-integrated lessons (Saubern, Urbach, koehler, & Phillips, 2020). Several research studies have used TPACK as a design framework for pre-service teachers’ ICT-integrated lesson design training (Ottenbreit-Leftwich, Glazewski, Newby, & Ertmer, 2010). Many teacher education programs have recognized the challenges associated with ICT-integrated lesson design using TPACK framework and proposed innovative teaching strategies (Angeli & Valanides, 2009). One of such strategies is collaborative lesson design (CLD) (Voogt et al., 2013). A large number of studies on collaborative learning have shown that formative peer feedback plays a critical role in enabling students to reflect on and continuously revise their work by receiving and providing concrete comments and suggestions on how to improve one another’s work (Yang, 2011).

Formative peer feedback aims to help peers improve their work and performance which involves reflective engagement (Falchikov & Blythman, 2001). The effects of formative peer feedback have been substantially evidenced to vary with the type of feedback (Wisniewski, Zierer, & Hattie, 2020). Generally, there are two types of formative peer feedback: cognitive feedback and affective feedback (Nelson & Schunn, 2009). Cognitive feedback involves summarizing, specifying and explaining aspects of the work under review (Chen, Hwang, Lai, & Wang, 2020; Huismann, Saab, van Driel, & van den Broek, 2018; Veerman & Veldhuis-Diermanse, 2001). Cognitive feedback can be classified into three sub-categories: scaffolding, evaluation, and questioning (Hoey, 2017). Scaffolding type of feedback aims to provide suggestions to the identified problems in peers’ work. Evaluation type...
of feedback plays an evaluative role and aims to make decisions and assess the quality of peers’ work. Questioning type of feedback aims to identify the problems and issue in peers’ work. Feedback that provides suggestions (i.e., scaffolding) and identifies problems (i.e., questioning) are especially effective by facilitating students’ interaction and knowledge construction (Kwon, Park, Shin, & Chang, 2019; Nelson & Schunn 2009). A recent study found that scaffolding was positively related to both learners’ perception of the formative peer feedback, and their willingness to improve based upon the feedback, but was not directly related with their performance improved in academic writing (Huisman et al., 2018).

Affective feedback uses affective language to bestow praise (“well done”) and criticism (“bad work”), or uses emoticons to convey emotion in text (Lu & Law, 2012). Affective feedback can be divided into two types: positive and negative affective feedback. Positive affective feedback is usually recommended and is found to be one of the most common features presented in the formative peer feedback (Cho, Schunn, & Charney, 2006). There are many studies on teachers’ praise of students (Kwon et al., 2019), but few studies examined the praise given by peers. Furthermore, there are mixed results on the effect of positive affective feedback on students’ learning improvement. Some studies showed the benefits of positive affective feedback by improving student’s learning (Duijnhouwer, 2010) whereas other studies found that positive affective feedback was ineffective for improving students’ learning, especially when the required task is cognitively demanding (Kwon et al., 2019).

One research gap identified from the past research is that they examined different types of formative peer feedback without identifying the specific level of feedback, which might explain the mixed results on the effect of formative peer feedback. The specificity of the feedback concerns the contextual relevance and the precise and appropriate amount of information provided in the feedback. Researchers argued that more specific feedback was more effective than general feedback for students’ learning (Chen et al., 2020; Ferris, 1997). It is assumed that the level of specificity of the feedback moderate the effects of the type of formative peer feedback in learning improvement. Taken together, the effects of formative peer feedback on improvement of work vary with the types of feedback as well as the levels of specificity of task-related information provided in the formative peer feedback.

There is hardly any research in the literature investigating how formative peer feedback affect pre-service teachers’ improvement on CLD with TPACK. This study examines the relationship between the type and the level of specificity of the formative peer feedback and the improvement of the ICT integrated lesson design for pre-service teachers. This study has three research questions: (1) Does formative peer feedback improve the quality of pre-service teachers’ ICT integrated lesson design? (2) Is the type of formative peer feedback related to the improvement of pre-service teachers’ ICT-integrated lesson design? (3) Does the level of specificity of the formative feedback moderate the relationship between the cognitive and affective peer feedback and the improvement of pre-service teachers’ ICT-integrated lesson design?

2. Method
2.1 Participants and learning context
40 pre-service Chinese language teachers (39 females) studying at the Nanyang Technological University were enrolled. Ten of them were aged from 20 to 25; 19 were aged from 26 to 30; and 11 were aged above 30. The lecturer had four years of experiences in both pre-service and in-service teacher professional development, and can use technology effectively for teaching and learning.

The learning task was to collaboratively design Technology-Enhanced Learning (TEL) for Chinese teaching and learning. The study was carried out in where each participant was equipped with a computer. The participants were grouped into 10 groups with 4 members. Before the CLD activity, the TPACK was introduced, the design principles for TEL and CLD, and the strategies of formative peer feedback. After the introduction, a 1.5-hour collaborative lesson design activity was conducted once a week for three weeks. The online platform used for participants’ CLD was Padlet, a commonly used online collaborating tool for Singapore teachers. In the CLD activity, the participants worked in groups to develop an ICT-integrated lesson design and posted it on the group space at Padlet (30 minutes). Then they visited other groups’ working space to provide formative feedback (30 minutes) using the commenting feature of the platform. After that they went back to own group space to improve the lesson design by addressing comments from other groups (30 minutes). Collaborative lesson design activities of week 2 and 3 were the same as week 1.
2.2 Data collection
As this study focuses on the effects of formative peer feedback on the improvement of pre-service teachers’ ICT-integrated lesson design, we collected the pre-service teachers’ ICT-integrated lesson design before the formative peer feedback and after the formative peer feedback, as well as the formative peer feedback provided and received. In total there were 60 lesson plans and 256 posted comments as formative peer feedback collected.

2.3 Research instruments
2.3.1 Coding of ICT-integrated lesson design
The unit of analysis of ICT-integrated lesson design is one group’s completed lesson design before and after the formative peer feedback activity. The quality of the ICT-integrated lesson design was coded using a TPACK coding scheme. TPACK is a well-acknowledged and dependable guide to evaluate the quality of TEL design (Koh, Chai, Benjamin, & Hong, 2015). The ICT-integrated lesson design was coded from seven TPACK domains as defined by Koehler and Mishra (2009). TPACK was measured by a scale where 10 being the lowest quality and 50 being the highest quality. Two trained coders coded the data. The inter-coder reliability was high (Cronbach’s alpha = 0.96).

2.3.2 Coding of formative peer feedback
The formative peer feedback was coded by adapting the existing peer assessment and feedback coding schemes (Chen et al., 2020; Veerman & Veldhuis-Diermanse, 2001). The unit of analysis was one commenting post. Each commenting post was categorized into cognitive and affective dimension. As mentioned, cognitive feedback consists of scaffolding, evaluation and questioning, and affective feedback includes negative affective feedback and positive affective feedback. Furthermore, each peer comment was coded on the level of specificity which consists of three levels: low, medium, and high level. The formative peer feedback was coded by two trained coders. In terms of coding of the type of feedback, the inter-coder reliability was moderate (Cohen’s Kappa = 0.543). With regards to the level of specificity, the inter-coder reliability test results indicted moderate and substantial agreement (Cohen’s Kappa = 0.576 for positive affective feedback, 0.681 for scaffolding, 0.544 for evaluation, and 0.571 for questioning, respectively).

3. Results
3.1 Quality of ICT-integrated lesson design, before and after the formative peer feedback
To address the first research question, Wilcoxon test was applied to investigate whether the quality of ICT-integrated lesson design improved after the formative peer feedback as data did not follow normal distribution. We found that the quality of lesson design after the formative peer feedback ($M = 13.76$, $SD = 15.00$) was significantly higher than that before the formative peer feedback ($M = 10.23$, $SD = 15.30$) ($Z = 2.82$, $p < .05$).

3.2 Type and specificity level of the formative peer feedback
Table 1 shows the number and percentage of formative peer feedback by type and level of specificity. It is found that there was no negative affective feedback in the formative peer feedback provided and received. 48% of the formative peer feedback were at low level of specificity, while feedback with medium level specificity accounted for 32% of all feedback, followed by feedback with high level specificity (21%).

<table>
<thead>
<tr>
<th>Category</th>
<th>Low level (Numbers, &amp; %)</th>
<th>Medium Level (Numbers, &amp; %)</th>
<th>High level (Numbers, &amp; %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive affective feedback</td>
<td>130 (64%)</td>
<td>35 (17%)</td>
<td>38 (19%)</td>
</tr>
<tr>
<td>Cognitive feedback</td>
<td>156 (39%)</td>
<td>158 (40%)</td>
<td>83 (21%)</td>
</tr>
<tr>
<td>Scaffolding</td>
<td>57 (32%)</td>
<td>82 (46%)</td>
<td>40 (22%)</td>
</tr>
<tr>
<td>Evaluation</td>
<td>23 (40%)</td>
<td>10 (17%)</td>
<td>25 (43%)</td>
</tr>
<tr>
<td>Questioning</td>
<td>76 (48%)</td>
<td>66 (41%)</td>
<td>18 (11%)</td>
</tr>
</tbody>
</table>

Note: Cognitive feedback is classified into three categories: Scaffolding, Evaluation, and Questioning.

3.3 Relationship between peer feedback and improvement of ICT-integrated lesson design
3.3.1 Correlation analyses
Correlation analyses were employed to address the second and third research questions. We found that the relationship between formative peer feedback and the improvement of ICT-integrated lesson design was moderated by both the type and the specificity level of the formative peer feedback. Specifically,
low level specificity positive affective feedback was negatively related to the improvement of the ICT-integrated lesson design, whereas medium level specificity positive affective feedback was positively related (r = -0.44, p < .05; r = 0.44, p < .05, respectively). Low level specificity evaluation cognitive feedback was positively related to the improvement of the lesson design (r = 0.40, p < .05). Medium level specificity questioning cognitive feedback was positively related to the improvement of the ICT-integrated lesson design (r = 0.48, p < .05).

3.3.2 Stepwise regression analyses

To further investigate how the type and specificity level of formative peer feedback affected the improvement of ICT-integrated lesson design, stepwise regression analyses were conducted to identity the significant predictors for the improvement of ICT-integrated lesson design (i.e., TPACK). A total of three specificity levels of formative peer feedback were included in the regression model. Regression analyses results showed that questioning with medium level specificity positively predicted the improvement of TPACK of the lesson design (β = 0.99∗∗, p < .01), whereas the high level specificity negatively predicted the improvement (β = -0.54∗, p < .05). Scaffolding type of cognitive feedback with medium level specificity negatively predicted the improvement of lesson design (β = -0.58∗, p < .05).

4. Discussion and conclusion

This study investigates the relationship between formative peer feedback and pre-service teachers’ improvement on ICT-integrated lesson design. This study found that the quality of ICT-integrated lesson design was significantly improved after the formative peer feedback. Furthermore, the relationship between formative peer feedback and the improvement of ICT-integrated lesson design was moderated by both the type and the specificity level of the feedback. The results suggest that questionings type of cognitive feedback with explanation was the contributing factor of the improvement of ICT-integrated lesson design; few numbers of scaffoldings with explanation helped improve pre-service teachers’ ICT-integrated lesson design. The findings confirm the role of formative peer feedback in improving pre-service teachers’ CLD in an authentic classroom setting.

Surprisingly, scaffolding type of cognitive feedback negatively predicted pre-service teachers’ improvement on ICT-integrated lesson design, in particular, peers’ scaffolding with explanation was related to the improvement of lesson design. One possible explanation on the negative effect of peers’ scaffolding type of cognitive feedback is that scaffolding was intended to provide suggestions to the identified problems in peers’ work or performance (Chen et al., 2020; Huismans et al., 2018; Veerman & Veldhuis-Diermanse, 2001). Therefore, based on scaffolding type of cognitive feedback provided by others, pre-service teachers could directly get the suggestions to the identified problems in peers’ work or performance without deeply think how to improve their lesson design. For example, one scaffolding type of cognitive feedback example was “Please provide explanations on words/Chinese phonetic alphabet, and provide scaffolding on given pictures.” With this feedback, the group knew clearly how to improve their lesson design.

Regarding positive affective feedback, feedback with low level specificity negatively related the improvement of ICT-integrated lesson design whereas feedback with medium level specificity positively related the improvement of the lesson design. Some studies have shown that compared with students not receiving formative peer feedback, the learners’ performance was not improved when they received pure positive affective feedback (Kwon et al., 2019). For example, one positive affective feedback with low level specificity was “This is very interesting.” The feedback did not include any concrete information related to the lesson design itself. Therefore, on one hand pre-service teachers might not get any concrete idea on how to improve their lesson design. On the other hand, they might consider their lesson design as perfect work, which did not need further revision. Unlike the feedback with the low level specificity, the positive affective feedback with medium level specificity included information about their lesson design on top of expressing positive affection. For example, one positive affection feedback with medium level specificity was that “This is good to combine students’ prior experience.” The group received this feedback further elaborated the part of individualized teaching in their lesson design according the feedback.

Regarding evaluation type of feedback, the level of specificity of feedback did not predict the improvement of the ICT-integrated lesson design. One possible explanation is that the frequency of evaluation type of cognitive feedback was relatively less than other types of feedback. In this study the number of peer evaluation type of feedback was 58 which was less than 10% of the total number of formative peer feedback (600). As a consequence, the evaluation type of cognitive feedback did not significantly affect the improvement of the ICT-integrated lesson design.
As questioning type of cognitive feedback, this study found that questioning with explanation was a significant predictor for the improvement of the ICT-integrated lesson design. This result echoes previous studies’ findings on formative peer feedback (Kwon et al., 2019; Nelson & Schunn, 2009). Several studies highlighted questioning as a key strategy in promoting students’ improvement (Kwon et al., 2019; Zhang, Lundeberg, McConnell, Koehler, & Eberhardt, 2010). For example, Zhang et al. (2010) found that soliciting idea and reframing idea questions helped students initiate their thinking, clarifying idea questions provided the students with opportunities to elucidate their thinking, and pushing for elaboration and checking for interpretation questions deepened the students’ thinking. Interestingly, this study also found that peers’ questioning with lots of explanations negatively predicted the improvement of the lesson design. The theory of zone of proximal development (ZDP) developed by Vygotsky (1929) provided a possible explanation for this result. ZDP refers to the distance between the actual developmental level as determined by individual problem-solving and the level of potential development as determined through problem-solving under guidance or in collaboration with more capable peers (Bruner, 1984). According to the theory of ZDP, peers’ questioning type of feedback is effective only when it is higher than pre-service teachers’ actual developmental level, and be lower than the level of potential development. When questioning type of feedback was too simple or too complicated, it might be either lower than pre-service teachers’ actual developmental level or higher than their potential developmental level in ICT-integrated lesson design. Therefore, peers’ less and too specific questionings did not facilitate pre-service teachers’ improvement in ICT-integrated lesson design, and such questioning type of feedback might even hinder their improvement in lesson design.

One limitation of the study was that the series statistical analyses used may not be able to explain the mechanism of how the lesson design was improved. More qualitative studies (e.g., uptake analysis, transitivity analysis) are needed to examine how pre-service teachers improve the lesson design through the formative peer feedback (van de Pol, Mercer, & Volman, 2019).

Despite the above limitations, the findings of this study contribute to the literature on pre-service teachers training of ICT-integrated lesson design and formative peer feedback for learning and improvement. The power of formative peer feedback to improve students’ performance or work has been recognized in earlier research (Gikandi & Morrow, 2016), but there has been little research on how formative peer feedback affect pre-service teachers’ improvement on ICT-integrated lesson design, and how the types as well as specificity of co-design related information moderated the effects of formative peer feedback. This study is the first to examine formative peer feedback with different types and levels of specificity.

In conclusion, this study found that formative peer feedback helps pre-service teachers’ improvement on ICT-integrated lesson design and this effect varied both with the types as well as with the levels of specificity of task-related information of the feedback. In specific, positive affective feedback and questioning with explanation type of cognitive feedback facilitated pre-service teachers’ improvement in lesson design. On the contrary, scaffolding with explanation hindered pre-service teachers’ improvement on ICT-integrated lesson design. The results inform how formative peer feedback affect pre-service teachers’ improvement ICT-integrated lesson design. This study has practical implications for educational practices in collaborative lesson design: peers are encouraged to generate more positive affective feedback and questioning type of feedback with explanations.

Acknowledgements

We would like to thank all the people who prepared and revised previous versions of this document. This study was funded by the Education Research Funding Programme, National Institute of Education (NIE), Nanyang Technological University, Singapore, project no. OER 17/19 CWL. The views expressed in this paper are the author’s and do not necessarily represent the views of the host institution.

References


Computational Thinking Activities in Number Patterns: A Study in a Singapore Secondary School

Shiau-Wei CHAN*, Chee-Kit LOOI*, Weng Kin HO*, Wendy HUANG*, Peter SEOW*, Longkai WU* & Mi Song KIMb

*National Institute of Education, Nanyang Technological University, Singapore
bUniversity of Western Ontario, Canada
*shiauwei.chan@nie.edu.sg

Abstract: In recent decades, there is a growing body of literature about CT and mathematics education. Thus, the purpose of this study was to identify the effectiveness of the integration of CT into mathematics classrooms in a Singapore secondary school, particularly in the topic of number patterns. A quasi-experimental design was employed in this study. 106 lower secondary students were involved in this study. 70 of them were assigned into the experimental group, while 36 of them were assigned into the control group. The students in the experimental group were given intervention unplugged Math+C activities and plugged Math+C activities using a spreadsheet. Meanwhile, the students in the control group received no intervention and were involved in traditional instruction. Both groups were given a pretest before the instruction and posttest after the instruction. The data obtained were analyzed using a two-way mixed-design ANOVA analysis. The findings revealed that there was a significant main effect of the pretest and posttest between the students from the experimental group and control group, but not a significant main effect of groups. Also, there was not a significant interaction between tests and groups. This study contributes to the area of integration of CT and mathematics in the instruction.

Keywords: Computational thinking, number patterns, secondary school, Math+C, quasi-experiment

1. Introduction

CT has been recognized as a collection of skills and understandings essential for new generations to be capable of using tools, but also able to understand the implication of their competences and limitations (Magana & Silva Coutinho, 2016). In 2014, Singapore launched the Smart Nation initiative, which has among its goals, developing its citizens’ digital literacy through coding or CT education for pre-schoolers to adults. In primary school, 10-hour coding classes have been made compulsory in the primary grades. Computing subjects are offered as part of the formal curriculum in secondary schools and junior colleges. The Infocomm Media Development Authority (IMDA) has launched several informal programs to develop CT skills and coding abilities including the Code for Fun Enrichment Programme, PlayMaker program, and Digital Maker program. The Science Centre also provides similar programmes to the general public.

Nowadays, there has been a rising interest in introducing CT into the mathematics classroom. Ho et al. (2017) reported on lessons designed to enrich science and mathematics through the integration of CT. In this study, Math+C is defined as the integration of Math and CT in the design and enactment of lessons. Thus far, the term CT has no universally accepted definition as a field by itself or as a field in the different disciplines like mathematics. In our present context, we define CT practices in the mathematics as involving decomposition (is the process by which the mathematics problem is broken down into smaller sub-problems or sub-tasks), pattern recognition (the action of looking out for common patterns, trends, characteristics or regularities in data), abstraction (the process of formulating the general principles that generate these recognized patterns), and algorithm design (the development of a precise step-by-step recipe or instructions for solving the problem at hand as well as a problem similar to it).
Various studies using experimental or quasi-experimental design methodologies have suggested positive correlations between CT and mathematics learning performance. Calao et al.’s (2015) study found that by using a Scratch visual programming environment to develop computational thinking, students in the experimental group can improve their performance in the mathematical process of reasoning, modeling, reasoning, exercising, and problem-solving. The study of Fidelis Costa, Sampaio Campos, & Serey Guerrero (2017) claimed that the development of math questions that are more aligned with CT can have a positive impact on students' problem-solving ability. These studies provide more evidence for the development of CT in mathematics education as a mechanism to promote students' capabilities. Hence, it is hypothesized that the quasi-experimental integration of CT and mathematics did in this study would have a positive impact on the learning performance of secondary students in Singapore. Thus, this study aims to investigate the impact of Math+C activities on the performance of secondary students in a Singapore school. There were two research questions being examined in this study: (1) Is there a statistically significant main effect of the pretest and posttest between the experimental and control group? (2) Is there a statistically significant interaction between tests (pretest and posttest) and groups (experimental group and control group)?

2. Method

2.1 Participants

This study employed a quasi-experimental design that non-randomly assigned students into the experimental group and control group. A total of 106 Secondary One students from a Singapore secondary government school involved in this study where 70 of them in the experimental group and 36 of them in the control group. There were 37 males and 33 females in the experimental group, as well as 20 males and 16 females in the control group. They came from three intact classes. The students already had the prior knowledge on how to evaluate an algebraic expression given an expression and an unknown, as well as on how to solve linear equations with one variable. The experimental group took part in the intervention with unplugged Math+C activities and plugged Math+C activities using a spreadsheet. Meanwhile, the control group was given traditional instruction. Nonetheless, both groups were administered with the pretest before the instruction and posttest after the instruction.

2.2 Instrument

The pretest and posttest were constructed by the researchers. Each test consisted of eight items as revealed in Table 1. The topic covered in both tests was Number Patterns, which was Chapter 7 in the Singapore Secondary Mathematics curriculum. There were three types of number sequences used in this study, i.e. arithmetic sequence, quadratic sequence, and geometric sequence. There were two categories of number sequences presented to students, namely numeral and figural. The validity and reliability of the instruments were determined by using the Rasch model. The construct validity for both tests was good as the values of raw variance explained by measures were greater than 40% (Linacre, 2012). Meanwhile, the reliability for both tests was high, i.e. more than 0.90 (Qiao, Abu, & Kamal, 2013).

Table 1. Test Items

<table>
<thead>
<tr>
<th>Type</th>
<th>Skill Tested</th>
<th>Sample Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arithmetic sequence (Numeral)</td>
<td>Identify terms of simple number sequences when given the initial terms</td>
<td>Fill in the missing terms in the following sequence: 7, ____, -5, -11, ____, -23</td>
</tr>
<tr>
<td>Quadratic sequence (Numeral)</td>
<td>Identify terms of quadratic sequences when given the rule</td>
<td>Find the 9th term of the sequence: 1, 3, 6, 10, …</td>
</tr>
<tr>
<td>Geometric sequence (Numeral)</td>
<td>Identify terms of geometric sequences when given the rule</td>
<td>Write down the next two terms of the number sequence: 9, 27, 81, 243, ____ , ____</td>
</tr>
</tbody>
</table>
Arithmetic sequence (Figural task with successive configurations)

(a) Use visual cues established directly from the structure of configurations to illustrate the pattern.
(b) Identify terms of arithmetic sequences when given the rule.
(c) Generate the rule of a pattern.
(d) Obtain an unknown input value when given the formula and an output value.

The diagram below shows a sequence of bricks.

Figure 1
Figure 2
Figure 3

(a) Draw Figure 4.
(b) Complete the following table.

<table>
<thead>
<tr>
<th>Figure Number</th>
<th>Number of bricks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

(c) Is there a Figure Number in the sequence that contains 136 bricks? Justify your answer.

Arithmetic sequence (Numeral)

(a) Use only cues established from any pattern when listed as a sequence of numbers or tabulated in a table.
(b) Generate the rule of a pattern.

Consider the following number pattern:

\[
1 = 3 \times 1 - 2 \times 1 \\
8 = 3 \times 2^2 - 2 \times 2 \\
21 = 3 \times 3^2 - 2 \times 3 \\
40 = 3 \times 4^2 - 2 \times 4 \\
\ldots \\
341 = 3 \times n^2 - 2 \times n
\]

Write down the equation in the 6th line of the pattern.
Deduce the value of \( n \). Explain or show how you figured out.

2.3 Computational Thinking Activities for Experimental Group

For the experimental group, the students were given intervention with unplugged Math+C activities and plugged Math+C activities using a spreadsheet. Four CT practices, i.e. pattern recognition, decomposition, algorithm design, and abstraction were employed throughout the instruction. The Math+C worksheets used in the unplugged Math+C activities were designed in line with these four CT practices.

Table 2 displays the instructional design of the plugged Math+C activities using the spreadsheet. There are four design principles for the integration of CT into lessons of number pattern based on CT practices, namely the principle of complexity, the principle of data, the principle of mathematics, and the principle of computability (Ho et al., 2020). The first principle is the principle of complexity. The tasks related to the topic of learning digital patterns are very complex to be broken down into subtasks. If the task is routine and can be easily solved using simple and well-known methods, decomposition cannot be utilized well.

The second principle is the principle of data. This task should consist of quantifiable and observable data that can be utilized, transformed, processed, and stored. Furthermore, the principle of mathematics is regarded as the third principle. We have to determine whether the task causes a situation or problem that can be mathematized. Mathematics is a problem constructed using mathematical terms. It includes altering the actual problem context to a mathematical problem in a precise and abstract technique. The task ought to be formulated abstractly so that it can be reasonably reasoned, described, and represented. The last principle is the principle of computability. We need to ensure a task solution that can be performed on a computer through a finite process.
Table 2. Sample of instructional design for plugged Math+C activities using a spreadsheet

<table>
<thead>
<tr>
<th>CT Practices</th>
<th>Sample of Instructional Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algorithm design</td>
<td>The students utilized the spreadsheet to perform recursion method and attained the final product as shown in figure below.</td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="Algorithm design" /></td>
</tr>
<tr>
<td>Pattern recognition</td>
<td>The students drew a scatter plot to represent the data to reveal the patterns of the data as exhibited in figure below.</td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="Pattern recognition" /></td>
</tr>
<tr>
<td>Decomposition</td>
<td>The students were guided to do the decomposition by breaking the problem into two smaller problems:</td>
</tr>
<tr>
<td></td>
<td>Problem 1: What is the starting number? Where to key in?</td>
</tr>
<tr>
<td></td>
<td>Problem 2: How do I use the recursion method to generate the number pattern?</td>
</tr>
<tr>
<td>Abstraction</td>
<td>The students were asked to generate the general formula using abstraction method as demonstrated in figure below.</td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="Abstraction" /></td>
</tr>
</tbody>
</table>

2.4 Lessons for Control Group

The control group was instructed using a more traditional approach. Unlike the experimental group, the students in the control group did not utilize any technology to learn the topic of number patterns. Furthermore, the worksheets used for the control group were routine problems, which was different from the worksheets that employed in the experimental group. The teacher used two worksheets when she taught the students in the control group. The students were asked to find the next two terms in the number sequence. Besides, the students were given a general term for a sequence of numbers, and they were required to find the term for the sequence. They were guided on how to generate general terms from a sequence of numbers in a worksheet.

3 Findings

Raw data are not always linear, so they should be converted to linear measures before conducting the parametric statistical test such as t-tests and ANOVA. In this study, the researcher utilized the Rasch model analysis to convert the raw data into linear measures, which means the scores were transformed...
into an interval scale. The person measures obtained in the Rasch analysis were then employed in the two-way mixed-design ANOVA analysis. This was to avoid problems related to the nonlinearity of raw test data and rating scales (Boone, Staver, & Yale, 2014).

The mean scores of pretest for the experimental group and control group were identified to ensure the students from both groups had equivalent capabilities in answering the questions to get rid of threats and support research (Gay, Mills, & Airasian, 2011). As shown in Table 3, the experimental group had a mean pretest score of 1.2183 (SD = 0.7250), while the control group had a mean pretest score of 1.2172 (SD = 0.5316). This indicated that the capabilities of the students from both groups were identical. The mean posttest scores for the experimental group were 1.4941 (SD = 0.7819) and the mean posttest scores for the control group were 1.4769 (SD = 0.7221).

Table 3. Descriptive statistics of pretest and posttest

<table>
<thead>
<tr>
<th>Test</th>
<th>Experimental Group (N = 70)</th>
<th>Control Group (N = 36)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Pretest</td>
<td>1.2183</td>
<td>0.7250</td>
</tr>
<tr>
<td>Posttest</td>
<td>1.4941</td>
<td>0.7819</td>
</tr>
</tbody>
</table>

A two-way mixed-design ANOVA was conducted to address research questions 1 and 2. From the tests of within-subjects effects in Table 4, it was noticed that there was a significant main effect of the pretest and posttest between the students from the experimental group and control group with F (1, 104) = 8.511, p = 0.004 < 0.005. Nevertheless, there was not a significant main effect of groups by referring to tests of between-subjects effects in Table 4. This is because F (1, 104) = 0.006, p = 0.937 > 0.05. Furthermore, there was not a significant interaction between tests (pretest and posttest) and groups (experimental group and control group) with F (1, 104) = 0.008, p = 0.930 > 0.05 as revealed in table 4.

Table 4. The analysis results of two-way mixed design ANOVA

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>0.004</td>
<td>1</td>
<td>.620</td>
<td>.006</td>
<td>.937</td>
</tr>
<tr>
<td>Test</td>
<td>3.410</td>
<td>1.000</td>
<td>3.410</td>
<td>8.511</td>
<td>.004</td>
</tr>
<tr>
<td>Group*Test</td>
<td>.003</td>
<td>1</td>
<td>.003</td>
<td>.008</td>
<td>.930</td>
</tr>
<tr>
<td>Errors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSBS</td>
<td>64.921</td>
<td>104</td>
<td>.624</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSWS</td>
<td>41.667</td>
<td>104</td>
<td>.401</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>106</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. ***p < .001

4 Discussions and Conclusion

For research question one, the results showed that there was a significant main effect of the pretest and posttest between the students from the experimental group and control group based on tests of within-subjects effects. This indicated that the pre-test and post-test of the experimental group and the control group are significantly different. However, there was not a significant main effect of groups by referring to tests of between-subjects effects. This implied that both groups did not make significant improvements in the posttest. For research question two, there was not a significant interaction between tests (pretest and posttest) and groups (experimental group and control group). It means that the changes from pretest to posttest for the experimental group and control group were somewhat similar. In other words, the intervention with unplugged Math+C activities and plugged Math+C activities using a spreadsheet had not much impact on the students’ performance in mathematics.
The findings did not support the hypothesis that integrating CT in lessons can result in improved mathematics learning. It means that the results of this study were not consistent with the results of previous studies with a positive impact of CT on students’ understanding of mathematical knowledge such as Calao et al. (2015) and Fidelis Costa, Sampaio Campos, and Serey Guerrero (2017). The possible reason why the involvement of students in the CT activities was not effective might due to the short duration of the CT activities given to the students in the experiment group. There was only one session of plugged Math+C activities using a spreadsheet with a duration of less than one and a half hours and unplugged Math+C activities with a duration of fewer than two hours. The experimental students may only engage in certain aspects of the problem-solving process during the CT activities, but these aspects are not sufficient to reveal significant differences with the control students. A follow-up interview or future study ought to be conducted to find out if there is a warrant to suggest that CT elements of the Math+C treatment influenced student learning.

The results obtained added the empirical evidence to the literature review on the incorporation of CT in mathematics and served as a guideline for the researchers, instructors, and school administrators in planning and designing the Math+C lessons in the schools. However, there were some limitations in this study. The sample involved in this study may not be representative of the population of lower secondary students in Singapore. Hence, a larger sample ought to be included in further studies to have a better generalization of the results. Another limitation was the short period of the intervention. So, the duration of the intervention should be expanded in future studies to evaluate the long-term impact on the students’ learning gains in the CT concepts and mathematical knowledge. In future studies, it is suggested to expand this study with the respondents from different demographics such as grade level, gender, race, school, computational tools, and so forth.

Acknowledgements

This work was supported by National Institute of Education under Grant Number OER 10/18 LCK.

References


Prompting Learner-Learner Collaborative Learning for Deeper Interaction: Conversational Analysis Based on the ICAP Framework

Shigen SHIMOJO\textsuperscript{a*} & Yugo HAYASHI\textsuperscript{b}

\textsuperscript{a}Graduate School of Human Science, Ritsumeikan University, Japan
\textsuperscript{b}College of Comprehensive Psychology, Ritsumeikan University, Japan
*cp0013kr@ed.ritsumei.ac.jp

Abstract: There are many studies investigating the type of interactions that are important for the success of collaborative learning. Recently, studies have investigated the use of the Interactive-Constructive-Active-Passive (ICAP) framework to promote collaborative learning, to guide and prompt discussions and encourage wanted behaviors. A previous study investigating grounding and conflict prompts revealed that the target utterances of each prompt were facilitated. However, the analysis of how these prompts deepened the interaction process has not been examined sufficiently. Therefore, the current study reanalyzed utterances generated by dyads; a protocol analysis employing a coding scheme based on the ICAP framework was utilized. The results indicated that interactive utterance is fostered by presenting two prompts. In future studies, we plan to investigate the relationship between learning performance and the ICAP framework.

Keywords: Interactive, constructive, coordination, argumentation, facilitations

1. Introduction

In cognitive, collaborative is beneficial for learners in order to learn various topics (e.g. Miyake, 1986; Shirouzu, Miyake, & Masukawa, 2002). However, collaborative learning does not necessarily outperform individual learning. Therefore, researchers have investigated what types of interactions are important for success and provide support to learners through the use of a prompt and a computer-supported collaborative learning (CSCL) setting (Dillenbourg & Fischer, 2007). Recently, researchers have investigated adaptive support using Intelligent Tutoring System (ITS) in a CSCL setting (Rummel et al., 2008). Therefore, this study focuses on facilitation prompts that support a dyad by the third person. First, from the viewpoint of the Interactive-Constructive-Active-Passive (ICAP) framework, it is crucial to explain why the interaction process is important in collaborative learning. The effect of facilitation prompts is then described.

1.1 Interaction fostering collaborative learning

Many studies have shown that collaborative is better than individual (e.g. Miyake, 1986; Shirouzu et al., 2002). It is important to undertake constructive interaction that learners externalize individualized ideas and obtain different perspectives through collaborative learning activities (Shirouzu et al., 2002). The ICAP framework (Chi, 2009; Chi & Wiley, 2014) has classified these activities. Passive activities include receiving learning material. Active activities include physical manipulation toward learning material. Constructive activities include externalizing a new idea deepened the given information. Interactive activities are when learners discuss substantially and do not ignore the partner’s contribution. The ICAP framework shows that learning increases as learners are engage in deeper interaction that is from passive to active to constructive to interactive activities (Chi & Wiley, 2014). There are two evidence for this hypothesis. First, ICAP is hierarchical and deeper interaction activities include shallow...
interaction activities. Second, knowledge-change processes underly each activity. Consequently, a support leading from constructive to interactive activities is called for in order to facilitate a deeper interaction between dyads.

Recently, some studies investigated the effect of support by the third person, classifying learning activities based on the ICAP framework. Wiggins, Eddy, Grunspan, & Crowe (2017) quasi-experimentally investigated whether the ICAP framework predicts learning performance in STEM class (i.e., Science, Technology, Entrepreneurship, and Mathematics). As a result, the instructor indicated that learners make more effort to support interactive activities because of the need to ensure sufficient time. Tan (2018) investigated the effect of two cognitive scripts focused on common ground in collaborative learning. It is known that learners do not undertake interactive activities superficially, rejecting an individual idea and not integrating ideas because of not processing that sufficiently. Accordingly, ICAP existed in these studies (Wiggins et al., 2017; Tan, 2018), but interactive activities are not easy to generate by simply collaborating.

It is known that collaborative learning is not successful because it is difficult for learners to understand their partner’s perspectives; this is linked to egocentrism conflicting with different perspectives and learners are asked to experience grounding and division of roles as a success factor (Hayashi & Miwa, 2009). Also, learners have difficulty in disagreeing with partner’s idea. Because learners have supporting tendencies their idea or perspectives instead of disagreement (Nussbaum & Kardash, 2005). These problems occurred when learners collaborate with partners. Therefore, it is assumed that we need to foster leading from constructive activities to interactive activities in a collaborative learning setting. In collaborative learning, facilitation by a third person is helpful (i.e., teacher and system) to support learners. However, it is not easy to foster interactive activities by a third person. Therefore, it is a challenge for the teacher to foster interaction.

1.2 Support of interaction using facilitation prompts

ITS has investigated adaptive supports based on the state of learners for individual learning (Anderson, Corbett, Koedinger, & Pelletier, 1995). Researchers in the area of ITS has recently investigated supports for collaborative learning. (Walker, Rummel, Koedinger, 2014). For example, Walker et al. (2014) investigated effect of adaptive support through the use of prompts based on learners’ state. Also, one of the types of support involves collaborative learning by pedagogical conversational agents (PCA) that is an interactive facilitator using language. Tegos & Demetriadis (2017) investigated whether PCA intervention based on academically productive talk (APT) which fosters building on their prior knowledge facilitates reasoning. As a result, learners who are intervened by PCA based APT outperformed learners who are not intervened by that. Recently, our study (not yet published) developed two different prompts for two conversational activities (i.e., coordination and argumentation) and investigated the effect of prompt that facilitate coordination or argumentation. Coordination is utterance leading to success in collaborative such as grounding and reaching consensus. Argumentation is utterance leading to strength epistemic status and dispute the idea deeply such as challenging and concession. Therefore, to investigate the effect of facilitation by presenting prompt developed based on the protocol data, we compared grounding condition presenting prompt fostering coordination and conflict condition presenting prompt fostering argumentation to control condition. As a result, the grounding and conflict conditions outperformed the control condition on each utterance activity. Consequently, these studies (Tegos & Demetriadis, 2017, Walker et al., 2014) show the effect that prompting by using the third person to intervene facilitates collaborative learning process.

From the above, providing a support facilitation of utterances related coordination and argumentation could yield interactive utterances from constructive because of removing the difficulty of understanding partner’s perspective and disagreeing with partner from learners in collaborative learning. In a previous study of ours (not yet published), we examined the effect of present two facilitation prompts that fostered coordination or argumentation in collaborative learning. However, the study has not showed whether two prompts are benefit for leading learners from constructive to interactive activities. Therefore, we investigate to conduct conversational analysis based on the ICAP framework. The prompt that fosters coordination could facilitate interactive utterances because learners need to establish common ground and reach consensus with each other. In addition, the prompt that fosters argumentation could facilitate interactive utterances because learners need to claim their idea based on their partner’s idea and disagree with partner’ idea.
1.3 Purpose and hypothesis

This current study focuses on becoming form constructive utterances to the interactive utterance that is a collaborative learning process where the interaction is deeper in ICAP framework (Chi, 2009; Chi & Wylie, 2014). Taking this consideration, the purpose of this study is to investigate the effect of presenting two prompts fostering coordination or argumentation that could facilitate from constructive utterance to interactive utterances. H1 is the prompt which fosters coordination could facilitate interactive utterances. The evidence for H1 is that learners could engage in interactive utterances because they need to acknowledge partner’s utterance, agree and disagree with the partner for consensus by facilitating sustaining mutual understanding and reaching consensus. H2 is the prompt that fosters argumentation could facilitate interactive utterances. The evidence for H2 is that learners could engage in interactive utterances because they need to consider the partner’s idea by facilitating challenge and concession.

2. Method

2.1 Participants

In this paper, we reanalyzed the dataset of our study (not yet published). A total of 94 learners (31 males, 63 females) participated with the average age being 19.85 (SD = 1.44). Participants were randomly assigned to the control, grounding, or conflict condition. Dyads collaborated without a prompt in the control condition. In the grounding condition, dyads received a prompt that facilitated coordination. In the conflict condition, dyads received a prompt that facilitated argumentation.

2.2 Experimental material and system

The present study utilized text about attribution theory, specifically that based on success and failure. In order to understand attribution theory, dyads were shown an episode. This is related to the task that dyads conducted in the experimental task. In the experimental task, Cmap Tools (https://cmap.ihmc.us/) was used because dyads made one individual concept map and one collaborative concept map. Our study (not yet published) developed prompts based on the coding scheme from Meier, Spada, & Rummel (2007) and Asterhan & Schwarz (2009). Meier et al. (2007) broadly classified into five dimensions: (1) communication, (2) joint information processing, (3) coordination, (4) interpersonal relationships, and (5) motivation. The following are prompts (excluding motivation) used in the grounding condition: (1) “Talk with each other to understand their idea and confirm if the partner understands the content”, (2) “Ask a question that is not written to the partner about causal attribution”, (3) “First, explain to the partners. Next, build a concept map,”, (4) “Don’t talk unilaterally. Let us consider the partners.”. On the other hand, Asterhan & Schwarz (2009) was broadly classified with the following three dimensions: (1) non-argumentative moves, (2) non-dialectical arguments, (3) dialectical argument. The following are prompts used in the conflict condition: (1) “Ask your partner a question about an utterance that is unclear “, (2) “Not just listen to your partner’s idea or claim but agree with it “, (3) “Let’s describe opposite ideas against your partner’s idea or claim. Include evidence for this.”.

2.3 Procedure

The experiment consisted of the experimental task and the explanation of the concept map and how to make one. The experimental task included individual phase and collaboration phase. Firstly, the experimenter explained the concept map and dyads read learning text about attribution theory individually. Second, in the individual phase (10 minutes), dyads conducted the task (Weinberger & Fischer, 2006) applied the causal attribution of success and failure to the episode (10 minutes). Narrative of a student was why the student was anxious about a promotion. In this study, dyads were asked to build concept maps about causal attribution and create individually at that time. Finally, in the
collaboration phase (15 minutes), the dyad created a concept map collaboratively referring individual maps. They could not see each other because of monitors and communicate with orality offline.

2.4 Dependent variables

In this paper, a coding scheme was developed based on Chi (2009), Chi & Wiley (2014) and Wang, Yang, Wen, Koedinger, & Rosé (2015) to capture active, constructive, and interactive utterance. Table 1 shows a part of the coding scheme and the relevant definitions. In addition, items of utterance about concept maps were added (e.g., utterance about concept map; reflection of an idea on the concept map). To investigate the reliability of coding, we conducted a third-person coding based on previous study (Schneider & Pea, 2014). A second coder coded 20% of the data which was randomly selected from the pool of all utterances. Next, to investigate the reliability, we conducted Krippendorf’s alpha coefficient. The results show that the coder’s matching rate was 0.74, which indicates that the coding was reliable.

Table 1. Coding scheme about ICAP

<table>
<thead>
<tr>
<th>Item and definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Active:</strong> Repeat/Paraphrase</td>
<td>“Peter said that my score was bad.”</td>
</tr>
<tr>
<td>The learners simply repeat the text or conversation.</td>
<td></td>
</tr>
<tr>
<td><strong>Constructive:</strong> Justify or provide reasons:</td>
<td>“I think the effort is internal because of himself.”</td>
</tr>
<tr>
<td>The learners propose one’s own idea or hypothesis.</td>
<td></td>
</tr>
<tr>
<td><strong>Interactive:</strong> Reflect idea on concept map:</td>
<td>After justifying or providing reasons of Constructive, “Okay, I will write effort in internal.”</td>
</tr>
<tr>
<td>The learners agree with partner and reflect the result or understanding of partner’s concept map.</td>
<td></td>
</tr>
</tbody>
</table>

3. Result

3.1 Conversational activities in collaborative learning

We firstly investigated how dyads undertook interactive, constructive, active utterances during collaboration without prompt. For this test, a within-subjects ANOVA was conducted. As a result, there was a significant difference among interactive, constructive, active utterances ($F(2, 58) = 13.10, p < .001$, partial $\eta^2 = 0.31$). Therefore, multiple comparisons of the Shaffer method were performed, and the ratio of interactive utterances was significantly lower than active and constructive utterance ($p < .001$).

3.2 The ratio of interactive utterances during the task

Next, a between-subjects ANOVA was conducted to investigate H1 and H2. Figure 2 shows the comparison of control, grounding, and conflict conditions with respect to the ratio of interactive utterances among interactive, constructive, active utterances in each condition. As a result, there was a significant difference between conditions ($F(2, 84) = 135.95, p < .001$, partial $\eta^2 = 0.76$). Therefore, multiple comparisons in the Shaffer method were performed and the conflict condition and the grounding condition were significantly higher than the control condition ($p < .001$).

*Figure 2. The comparison of interactive of 3 condition. Also, error bar is standard deviation.*
In addition, the conflict condition was significantly higher than the other two conditions (p < .001). These results supported H1 and H2. However, the average ratio of no labeled utterance that is not active, constructive, or interactive was 0.81 in the conflict condition. Many of learners talked about how create a concept map. This result indicates that active, constructive, and interactive utterances depict a small ratio in conversation even if dyads in the conflict condition undertook such utterances.

4. Discussion

4.1 Interaction process of in collaboration learning based on the ICAP framework

The purpose of the present study was to investigate how the interaction process became deeper through the use of grounding prompts and conflict prompts. First, conversations were analyzed in learner-learner collaborative learning settings without prompts with a protocol analysis based on the ICAP framework. The results revealed that dyads use fewer interactive utterances in collaborative learning than in other activities of this study. As majority of research shows, learners do not demonstrate interactive utterances as necessary (e.g. Hayashi & Miwa, 2009). The present study presented evidence of these results by coding utterances based on the ICAP framework. However, this study chose to focus on utterances instead of all behaviors. Chi & Wylie (2014) showed that building concept maps is a constructive activity. Therefore, we will need to use behavior in order to investigate collaborative process in detail.

4.2 Prompting dyads to the deeper interaction process

The results of the comparison between conditions showed that the use of prompts facilitated interactive utterance. This indicates that interactive utterance is fostered by facilitating coordination or argumentation. Therefore, H1 and H2 were supported. Interactive utterance of the ICAP framework included activities that were equivalent to coordination and argumentation in this study. For example, communication in one of the coordination prompts leads to the construction of common ground which includes asking and answering to understand other contents of utterance (Clark & Schaefer, 1989). In addition, dialectical argumentative moves in one of the argumentative prompts the challenge and rebuttal that criticizes each other (Schwarz, Neuman, & Biezuner, 2010). Furthermore, dyads in the conflict condition had a higher ratio of interactive utterances than those in the grounding condition. Dyads in the conflict condition needed to take partner’s idea into consideration because they were not able to critique without the partner’s idea. Therefore, interaction becomes deeper by presenting a conflict prompt compared to a grounding prompt.

Another important point is that the coding schema developed in this study was based on Chi (2009), Chi & Wylie (2014) and Wang et al. (2015) and included interactive utterance type (e.g., Acknowledgement of partner’s distribution) and not utterance necessarily generated by presenting a conflict prompt. In other words, a conflict prompt facilitates more interactive utterances compared to a grounding prompt, but the result also showed utterances related ICAP were a small ratio of all conversations. In future studies, it is crucial to present the prompt that fosters interactive utterance directly and investigates what part of the interactive utterance is facilitated through conflict prompts.

5. Conclusion

To lead from constructive to interactive utterances, learners need to be facilitated coordination or argumentation. Our previous study investigated whether dyads who received prompts generate more coordination utterances and argumentation utterances. Therefore, this study investigated whether interactive utterances are generated through the use of prompts that foster one of the interactive utterances by reanalyzing our previous study utilizing a conversational analysis based on the ICAP framework. The results show that the ratio of interactive utterances was greater in grounding and conflict conditions than in the control condition. In addition, dyads presented with the conflict prompt generated more interactive utterances than dyads presented with the grounding prompt. However, dyads do not generate interactive utterance as a deep process sufficiently though some interactive utterances
are facilitated by presenting prompts in attempt to foster one of these processes. In addition, we need to investigate the relationship between learning performance and the interaction process in a future study. The present study contributes to CSCL study that aim to lead deep interaction of learners in collaborative learning by providing a facilitation prompt.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP20H04299.

References


Shimojo, S. & Hayashi, Y. (not yet published).


Towards the Design of a Robot Peer-Tutor to Help Children Learn Math Problem-Solving

Aaron Nol BAUTISTA*, Jabin Raymond GERARDO*, Harvey LALLAVE*, Patrick Luigi LATORRE*, Ethel ONG**, Jocelynn CU*, Minie Rose LAPINIDb & Auxencia LIMJAPc

*College of Computer Studies, De La Salle University, Philippines  
**Br. Andrew Gonzalez FSC College of Education, De La Salle University, Philippines  
*cJosé Rizal University, Philippines  
*ethel.ong@dlsu.edu.ph

Abstract: Collaborative learning allows learners to work together as peers to understand the given problem and to formulate different strategies for solving. With the rising popularity of social robots, education researchers have explored the varying roles these technologies can fulfill in interactive learning environments, such as tutors and learning companions. In this paper, we describe our work in investigating a robot’s potential as a peer-tutor for Grade 1 students learning math problem solving. Combining audio-visual cues in presenting addition and subtraction problems, our robot, named Vi, worked with 12 children age 6-8 years old as they try to understand and formulate a solution plan for the given problems, using a dialogue framework that leverages on the mathematical thinking process. Our results showed mixed feedback from children; some found Vi fun to talk to while others felt pressured by Vi’s questions specifically for those participants who are not fluent in the English language.

Keywords: peer tutor, collaborative learning, math problem solving

1. Introduction

Solving word problems for basic addition and subtraction operations are taught in primary level classrooms as early as the first grade (Limjap, 2011). Huge class size and rigid academic calendar, however, force learners to follow a set of problem solving procedures prescribed by their teachers. Furthermore, frequent drills and exercises that are essential to the development of mechanical skills in performing basic math operations are oftentimes done in isolation. These practices may not be meaningful for all types of learners and can cause disinterest and non-completion of the tasks. For learning to occur, students need to have a deep understanding of the concepts beyond simple application of the stated procedures (Michaelis & Mutlu, 2019; Pehkonen et al., 2013). They should be able to reason through the problem and to justify their thinking. These abilities can be developed as early as Grade 1 and can be transferred to other learning contexts (Pehkonen et al., 2013).

Previous studies have established the importance of student-teacher and student-student interactions in enhancing learning (Mercer & Sams, 2006). But when exercises are administered, a teacher alone or a parent with insufficient background in teaching math may encounter challenges in providing the support needed by learners. In such situations, robots can be used to augment the need for social interaction and peer learning as a strategy for developing math problem solving skills. Acting as peer-tutors, facilities afforded by a robot's verbal and non-verbal cues can be used to scaffold learners towards the acquisition of skills in problem solving (Michaelis & Mutlu, 2019). In this paper, we describe our peer-tutor robot, named Vi, that can work with Grade 1 children during math word problem solving.

2. Designing Vi, the Robot Peer-Tutor

Vi is developed on top of the NEC PaPeRo (Partner-type Personal Robot) robotic platform shown in Figure 1. A number of built-in physical facilities of PaPeRo were utilized to mimic non-verbal cues
typically exhibited in human-to-human conversations. LEDs in PaPeRo's mouth would light to indicate that Vi is speaking. LEDs in the cheeks and mouth change colors and patterns to indicate emotions such as smiling or blushing. Movements such as nodding and head-shaking mimic signs of approval or rejection of the learner’s answers. LEDs in the robot’s ears would light to indicate that Vi is listening. Vi can hear the learner's input through two microphones found at PaPeRo's forehead and back. The LED on the forehead would light when Vi is thinking and/or processing information. These cues are meant to provide unobtrusive feedback and enable Vi to take part in the exchange without disrupting the learner's train of thought (Johnson et al., 2000).

A second modality using a laptop is incorporated for the visual presentation of text and image depicting the objects described in the word problem, as shown in Figure 2. A chat log provides a trace of the dialogue between the learner and Vi. The dialogue follows the mathematical thinking process described in (Boonen et al., 2016; Limjap, 2011). It is comprised of a series of questions that guides the learner in understanding the problem, formulating the solution and counting the objects.

![Figure 1. NEC PaPeRo robotic platform (www.nec.com).](image)

![Figure 2. Visual presentation of text and image depicting the objects described in the word problem.](image)

Following the prescribed learning competencies for teaching mathematics to primary schoolchildren by the Department of Education (Limjap, 2011), three types of word problems are used: join-result unknown, separate change unknown, and compare where quantity is unknown. Word problems given to Grade 1 learners focus on whole numbers, and addition and subtraction operations.

3. Results and Discussion

Preliminary end-user validation was conducted with twelve (12) Grade 1 students to identify the robot's strong and weak attributes as a peer-tutor. The three types of story-based math problems are given in a single learning session. Two questions for each problem type are available; the learner is given at least three and at most six problems, depending on his/her performance in prior problems of similar type. Each learner worked individually and engaged in dialogue with Vi with an average session of 30 minutes to one hour.

Social interaction can stimulate interest and motivate the learner to perform the required task. Vi portrays the role of a peer by exhibiting social abilities that can motivate collaborative learning behavior. Various factors can affect the social interaction needed for collaborative learning, including Vi’s ability to build rapport with the learners, turn-taking strategies and relevance of Vi’s responses. Table 1 shows the average scores from children’s feedback of Vi’s social interaction skills on a 5-point Likert scale.

While 92% of the learners found Vi fun to talk to, one participant, S2, had a differing opinion. S2 interacted with Vi twice with both interactions being incomplete, thus leading to a negative experience. S3, S11 and S12 did not consider Vi as their friend. While S3 shared that he/she is afraid of robots, S11 and S12 were not able to build social rapport with Vi. For the other learners, it was observed that at the onset of the interaction, they tend to seek an approval from Vi whom they regarded as a tutor first before answering the questions. Upon gaining familiarity with the robot, the learners gradually started to respond to Vi on their own.

S3 also did not find Vi’s voice friendly. This could be attributed to the monotonous voice afforded by the text to speech functionality that felt more like a robot than a human. Peer-to-peer conversations, however, are characterized by spontaneity and variety. S12 viewed Vi as a tutor by
stating that the robot asked too many questions thus, making the experience less interesting. This echoed the findings in Chan & Ong (2018) where younger participants perceived the agent to serve the role of a tutor when it asked many questions.

Table 1. Results of feedback from learners on Vi’s social interaction abilities.

<table>
<thead>
<tr>
<th>Survey Item</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>I find Vi fun to talk to.</td>
<td>4.50</td>
</tr>
<tr>
<td>I consider Vi as my friend.</td>
<td>3.92</td>
</tr>
<tr>
<td>I find Vi’s voice friendly.</td>
<td>4.50</td>
</tr>
<tr>
<td>Vi’s replies are interesting.</td>
<td>4.25</td>
</tr>
</tbody>
</table>

4. Further Work

We presented our robot peer-tutor, Vi, that guides Grade 1 students in learning how to solve math word problems through a three-phase conversational flow: understanding the problem, formulating the solution and counting the object. Results from validation with children showed mixed sentiments. They find Vi fun to talk to as a peer, giving an average score of 4.50 out of 5. Three (3) learners failed to recognize Vi’s role as a peer due to their apprehension with robots, lack of rapport and the repetitive responses of Vi. The inability to carry on a smooth conversation due to technical challenges in speech recognition and synthesis caused difficulty for children to learn with Vi. Furthermore, the robot’s role as a tutor led it to generate continuous prompts as a necessary component of the mathematical thinking process. This, however, caused stress among the learners.

The work presented here described a single session validation with 12 children to provide preliminary insights on the robot’s potential as a peer-tutor to aid Grade 1 math learners. Assessing child-robot interaction in a learning context takes time (Kory-Westlund & Breazeal, 2019). There is a need to conduct experiments over a longer term, with repeated use spanning several weeks to generate better insights on how the robot is able to fulfill its roles as peer and tutor, and its significant contribution as a learning companion.

Acknowledgements

We would like to thank NEC Japan and NEC Philippines for lending us a PaPeRo robot to test its facilities and for providing the necessary technical support in working with the robot.

References


Changes in the effect of concept map-based autonomous learning under different levels of self-regulation

Lu CHENG*, Fan CHEN*, Jiayu NIU*, Xueying XU* & Ning MA*

*School of Educational Technology, Faculty of Education, Beijing Normal University, China
*harsening@bnu.edu.cn

Abstract: This research is to explore the changes of concept map-based autonomous learning effect and its correlating with self-perceptions, motivational beliefs and the use of learning strategies which are the components of self-regulated learning. A mixed approach is employed with quantitative data from a self-regulation strategies questionnaire and evaluation of concept map, as well as qualitative data from a follow-up interview with thirty-three postgraduate students participated in this research. Our results indicate that the changes in the effect can be divided into three traces according to different trends, spikes and stability. Although it doesn’t reveal the significant correlations of the level of self-regulation and changes in the effect, we find that students have perceived the factors that may affect their changes of learning performance. This study highlights the characteristics of the autonomous learning effect trend during a period of time, and contributes to the instructional use of concept map as well.

Keywords: Autonomous learning effects, Self-regulation, Concept map

1. Introduction

The focus of learning in schools gradually shifts from a knowledge-based to a skills-based curriculum to adapt to the high-speed social development (Oates, 2019). Accordingly, the ability of autonomous learning has attracted much attention. The gradual increase in the importance of autonomous learning places high demands on learners’ self-regulation which is an integrated term referring to their motivational beliefs, metacognitive skills, and use of learning strategies (Cleary, & Platten, 2013).

Researchers have proposed numerous instructional strategies for improving self-regulation, such as the use of technology to support learning (Schraw, Crippen, & Hartley, 2006). Concept maps are often used, which is a support tool to affect autonomous learning through enhancing self-regulation strategies to varying degrees (Stevenson, Hartmeyer, & Bentsen, 2017). Novak (1984) thought learners utilizing concept maps in autonomous learning can not only monitor the process of their learning but also assess its outcomes. However, the effects on learning with concept maps are questionable, particularly for low self-regulated learners because autonomous learning requires a high degree of self-regulation to succeed (Dabbagh, & Kitsantas, 2004). Therefore, we attempt to explore the continuous changes in the effect of concept map-based autonomous learning under different levels of self-regulation, with the following questions: (1) How does the effect of learners' autonomous learning change toward concept map-based activities over time? (2) What is the correlation between autonomous learners' self-regulation and changes of learning effect toward concept map-based activities? (3) By follow-up interviews, how are the changes in autonomous learning effect influenced by self-regulation?

2. Method and Results

This research involved 34 first-year postgraduate students (32 females) participating in the same 18-week specialized course where autonomous learning took the form of English literature reading and concept mapping. During the autonomous learning, students read materials one week before class, and completed the knowledge construction based on concept maps. The evaluation of concept map includes
concepts (1 point), relationships (5 points), hierarchies (5 points), cross links (10 points), examples (1 point) and branches (1 point or 3 points) (Markham, Mintzes, & Jones, 1994). After the curriculum, study adapted the Self-Regulation Strategy Inventory-Self Report (SRSI-SR) (Cleary, 2006) to examine the frequency with which students engage in various adaptive and maladaptive regulatory behaviors. The modified SRSI-SR instrument ($\alpha = 0.905$) is still a 28-item scale utilizing a 7-point Likert scale.

To answer the first question, a second-order clustering is used to identify the whole features of changes in 8 concept maps scores per student, dividing the students into three traces (see Figure 1).

![Figure 1. Performance of Student in Autonomous Learning](image)

To reveal the differences of changes in learning effects under different levels of self-regulation, students were divided into three groups according to the self-regulation scores by the rate of 27%, 46% and 27% (Kelly, 1939). Then, according to Fisher's exact test results, there was no significant difference ($p > 0.05$) between changes traces and self-regulation levels. To have a deeper understanding of the correlation, 4 typical interviewees were sampled from each cluster of the K-Modes clustering algorithm which presented the correlation between variables and similarity between samples (see Table 1).

<table>
<thead>
<tr>
<th>Clusters of learning behavior pattern</th>
<th>Cluster A (n = 22)</th>
<th>Cluster B (n = 6)</th>
<th>Cluster C (n = 6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR level</td>
<td>Level 2 (Middle)</td>
<td>Level 3 (High)</td>
<td>Level 1 (Low)</td>
</tr>
<tr>
<td>Change trace</td>
<td>1 (Low &amp; steady)</td>
<td>3 (High, slowly falling)</td>
<td>3 (High, slowly falling)</td>
</tr>
<tr>
<td>Student ID</td>
<td>6, 27</td>
<td>8</td>
<td>11</td>
</tr>
</tbody>
</table>

From the thematic analysis results, two themes were extracted: self-perception, and the role of self-regulation in autonomous learning (see Table 2). The learning method had positive effects on students with high learning motivation (e.g., 8), but increased academic burden for students with low learning motivation (e.g., 11). Regarding the role of self-regulation, we found that the scores of students' concept map aiming at completing the task (e.g., 6) were generally low but slowly rose, while students whose main goal was to acquire knowledge (e.g., 8) had a good and stable learning effect. During the task, three types of help seeking behavior were captured: actively seeking external tool (e.g., 8), passively waiting for help (e.g., 27), and none attempting. When external support was not enough, students constantly adjusted their learning strategies to improve the learning effect (e.g., 8). Moreover, we found that some students with high self-regulation level had poor autonomous learning effect, which could be because they gave up self-regulation.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Themes</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-perception for the learning method</td>
<td>Motivations for learning</td>
<td>8: I should complete some learning task but don’t do well, I will feel frustrated and failed to meet my requirements.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11: If I want to be motivated to learn, I need external stimulus.</td>
</tr>
<tr>
<td>The role of self-regulation in</td>
<td>Goal setting</td>
<td>6: My main task is to complete the concept map.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8: My learning goal is to understand the knowledge related to</td>
</tr>
</tbody>
</table>
3. Discussion

Although previous researches have showed that self-regulation is the best predictor of academic performance (Young, 1996), we recognized from the quantitative analysis that the correlation of self-regulation and continuous performance in knowledge construction was statistically insignificant. The interview analysis indicated that some of the main self-regulated components did have an impact, for learners with specific self-regulation level and change traces were accustomed to use different SRL strategies (i.e., goal setting and help seeking). It is noteworthy that learners with better self-regulation usually took mastery as their personal goal, had specific plans, and tend to be more persistent and less likely to burn out over the long terms, which is consistent with previous studies (Wilson, & Kim, 2016).

However, many other factors (e.g., task difficulty, task value) were neither controlled nor focused because of the realistic conditions. In addition, the SRRI-SR scale in this research mainly measured learners' SRL strategies, but didn't exactly include the dimension of motivation beliefs which resulted in a long-term response. In this study, learners who had strong external motivation but failed to internalize in time gradually abandoned self-regulation though they had a high level of self-regulation. Thus, it is encouraged that studies be conducted to examine the influence of these variables.

Furthermore, this research used concept map as a support and assessment tool, but there was no specific investigation on how learners received and used the feedback from it. Future study could consider to explore the progress of using concept map as a cognitive and metacognitive tool.

References

Online Collaborative Kit-Build Concept Map

Aryo PINANDITO\textsuperscript{a,b*}, Yusuke HAYASHI\textsuperscript{b} & Tsukasa HIRASHIMA\textsuperscript{a}
\textsuperscript{a}Department of Information Engineering, Hiroshima University, Japan
\textsuperscript{b}Information System Department, Computer Science Faculty, Universitas Brawijaya, Indonesia
\textsuperscript{*arya@ub.ac.id}

Abstract:
Many studies show promising benefits of using concept maps as an effective approach to support learning, thus enhance students' understanding. Concept mapping with the Kit-Build concept map framework is known as one practical approach to improving students' understanding of learning material. Supporting meaningful learning in an online collaborative learning environment is challenging, where the students need an effective and meaningful interaction during the collaboration. Both the use of concept maps and collaborative learning are promising to promote an engaging learning environment, thus yielding a better learning outcome. This research introduces a newly developed concept mapping tool that fuses collaborative learning into the Kit-Build concept map framework. It allows students to learn by collaboratively working with concept maps in a distant real-time learning environment. This research investigates the tool's effectiveness in a learning environment where the students use different concept mapping approaches to learn with concept maps online collaboratively.

Keywords: collaboration, concept map, Kit-Build, learning effect, online

1. Introduction

A learning framework, namely Kit-Build concept-map framework, is a closed-ended concept mapping approach where the students reconstruct a concept map from the provided concept and link components of a teacher's concept map (Hirashima, Yamasaki, Fukuda, & Funaoi, 2015). In a learning environment with the Kit-Build concept map framework, students learn and express their understanding by reconstructing the teacher's concept map components. In its practical use, the teachers and the students use a computer-supported concept map authoring tool to create and reconstruct the concept maps, respectively. Several studies related to the Kit-Build concept map framework show that the use of Kit-Build concept map in learning context has many advantages in improving the students' understanding of learning material (Alkhateeb, Hayashi, Rajab, & Hirashima, 2015; Andoko, Hayashi, Hirashima, & Asri, 2020). The use of concept mapping tool may also provide some assistance in the concept map creation activities (Pinandito, Az-zahra, Hirashima, Hayashi, 2019).

Collaborative learning is known to promote greater benefits towards students' self-discovery, cognitive achievement, and productivity while engaging them to an active knowledge-sharing, social, and psychological learning environment (Laal & Ghodsi, 2012; Ibrahim, Shak, Mohd, Ismail, Perumal, Zaidi, & Yasin, 2015). As education systems were shifting from offline classroom to a distant or online context, the adoption of online education technology raises significantly. Activities in distant learning become more challenging, especially in learning activities that involve collaborative work.

This research introduces a new approach to learning with the Kit-Build concept map that allows students to learn, discuss, and work collaboratively in the distance and in real-time through an online concept mapping tool. Even though a similar tool has been developed (Farrokhnia, Pijeira-Diaz, Noroozi, & Hatami, 2019), this research has its implementations that extend its previous functionalities to the extent that the students and teachers can discuss and collaboratively work in real-time. The extension removes the offline barrier of the current concept mapping tool that supports a meaningful interaction among learners when they learn with Kit-Build concept maps (Wunmasri, Pailai, Hayashi, & Hirashima, 2018; Sadita, Hirashima, Hayashi, Furtado, Junus, & Santoso, 2019; Prasetya, Widiyaningtyas, Hirashima, & Hayashi, 2019). Furthermore, the effectiveness of two concept mapping approaches in improving students' understanding of a learning material while using the online collaborative concept mapping tool is also investigated.
2. Methodology

An experiment is designed to evaluate the tool's effectiveness in the context of an online real-time collaborative learning that uses the Kit-Build concept map method. The experiment involves 40 international graduate students divided into two groups of dyads, i.e., Collaborative Scratch Mapping (CSM), and Collaborative Kit-Building (CKB). The dyads in the CSM group create a concept map from scratch, while dyads in the CKB group create the concept map from a pre-defined Kit-Build concept map component (kit). An English learning material is prepared as the learning subject for the students to learn and comprehend its contents.

For the experiment, an online concept mapping system has been specifically built following the experiment flow. All participants were asked to represent their understanding in the form of a concept map while also discussing the map they made. Every participant uses a separate computer and rooms in such a way that they neither can see nor have a face-to-face discussion. Hence, simulating an online distant learning activity where direct verbal discussions are difficult or impossible to conduct. As shown in Figure 1, all participants were given training about how to use the tool, read the learning material, and answer some pre-test questions before doing the concept mapping with their partner collaboratively. The post-test and the delayed-test measure the students' understanding after they collaborate with their partner and measuring their knowledge retention, respectively.

![Figure 2. Experiment flow.](image)

3. Results

According to the analysis and comparison of the pre-test result, as shown in Figure 2, both the CSM and CKB groups have a similar level of understanding. According to the students' post-test results, learning collaboratively with the Kit-Build concept map is shown effective in improving the students' understanding of learning material. Their understanding improves significantly, according to the Wilcoxon paired test ($p$-value < 0.001). Furthermore, it can also be seen that their understanding is starting to decay after several days. The decline in their understanding is statistically significant as per Wilcoxon paired test $p$-value is less than 0.001. The comparison of the students' test scores from both groups is shown in Figure 2.

![Figure 2. Test score comparison between two concept mapping approaches.](image)
4. Conclusion and Future Works

The concept mapping tool introduced in this research, which combines the Kit-Build and collaborative learning approach, is promising to use in an online collaborative learning environment that uses concept maps. The online concept mapping tool introduces a new approach to learning with Kit-Build concept maps that combine the Kit-Build concept mapping activities with a collaborative learning environment. It creates an integrated online collaborative learning environment that incorporates concept mapping activities. The students can use the tool to collaboratively create concept maps with their partners and make an effective discussion that helps them learn while creating or reconstructing a concept map. The online collaborative learning with the online collaborative Kit-Build concept map tool also improves students' understanding of a given learning material regardless of the concept mapping approach used during concept mapping activities. The collaborative concept mapping activities with the Kit-Build concept map can now be conducted online in real-time where distance learning is an issue.

Still, there are many research opportunities to be addressed as the future works of this research. Further analysis of the resulting concept maps, analyzing the students' activities and their discussions during the collaborative work, and evaluating the usability of the tool in practical use are only several interesting future research topics that will benefit this research more.

Acknowledgements

This research is partially supported by JSPS KAKENHI Grant Numbers 19H04227.

References


Proposal of Note-map for Collaborative Reading Using an E-book System

Juan ZHOU*, Hideyuki TAKADA & Chengjiu YIN
a School of Environment and Society, Tokyo Institute of Technology
b College of Information Science & Engineering, Ritsumeikan University, Japan
c Information Science and Technology Center, Kobe University, Japan
*juan.z.kt@gmail.com

Abstract: E-books are continually being introduced to educational institutions. Many e-books systems have a note-taking function to support learners’ record of their reading by themselves. However, collaborative note-taking provides many benefits to learners, and as a way to improve reading efficiency and stimulate students’ learning motivation, collaborative reading has been widely used in educational fields. In this study, we describe a note-map function to support students’ collaborative reading. Students share their notes, edit others’ notes, communicate with others using the note-map function, and a scenario of using the system is presented.

Keywords: Collaborative reading, note-taking, E-book system, online learning

1. Introduction

In the last decade, there are a lot of online learning platforms developed for online teaching and learning. Especially recently, online learning attracts attention from many educational institutions. Enriching the online educational system and improving the online educational method were required since the past.

In Japan, e-books are continually being introduced to educational institutions and many e-book systems have a note-taking function. In our previous research, we found that the students took notes even if the teacher did not require it (Yin et al., 2015; Yin et al., 2017; Zhou, Takada, & Yin, 2019). Taking notes helps students to understand the materials and organize their way of thinking and encourages students to build connections between what is presented and what they already know (Peper, & Mayer, 1986).

After taking notes, they always share their notes to learn from each other. Collaborative learning is an effective formal education strategy, as Vygotsky’s ideas concerning the zone of proximal development provide strong support for the inclusion of collaborative learning strategies in classroom instruction (Doolittle, 1995). Al-Zaidi, Joy, & Jane (2013) indicated that social network sites have been used to share notes during lectures. They found that applications with social interaction features would enhance students’ current note-taking practice during lectures. Collaborative note-taking provides important benefits: better learner engagement, collaborative learning, and knowledge building (Silvestre, Vidal, & Broisin, 2014).

In this study, the authors propose a note-map function in the e-book system to support students to review their reading path, share their notes, edit others’ notes, communicate with others, and present scenario of using the system.

2. Relevant Literature

There are some social learning platforms for students’ collaborative reading (e.g., Perusall, WiREAD) that aim to change the reading experience from solitary to collective (King, 2016).

Silvestre, Vidal, & Broisin (2014) developed Tsap-Notes, an open micro-blogging platform dedicated to collaborative note-taking that can be used as a standalone application, or fully integrated into existing virtual learning environments. The Tsap-Notes was useful when the time of preparing
exams has come. Reilly, & Shen, (2011) applied a student-centered collaborative learning pedagogy into the lecture environment through a novel Smartphone-based real-time collaborative note-taking application – GroupNotes – that encourages students to proactively engage themselves by means of student-student interaction in a lecture. Popescu, E., et al. (2016) proposed a mobile application specifically conceived for educational settings, called EduNotes. Students can write notes associated with a specific lecture slide and share them with the peer.

3. Method

3.1 DITel System

We have developed the E-book system DITel to support students’ reading and collocate log data in the class (Yin & Hwang, 2018). Students read it by clicking “Prev” and “Next” buttons and write notes, highlight, underline, bookmark a page (Figure 1). As previous research, even if note-taking was not required, many students took notes and made highlights during their study (Zhou, Takada, & Yin, 2019).

3.2 Note-map

We designed note-map including students’ reading paths to support students with their reflection on their reading behavior and rethinking their reading by looking at others’ reading paths. This function visualizes the students’ reading paths to help their review. We expect this function to enable students’ reflection by their reading paths.

Students decide on sharing their note or not when they are taking the note. Students click the “Share” button if they want to share their notes with others. The notes will be shown in the note map. Note map includes pages, reading trajectories, and notes. In Figure 2, the yellow dot represents a page, and the gray arrows represent the movement of direction. The orange rectangle shows the notes on the page. All users can read their note map to review and read others’ note map to learn from others.

3.3 Scenario

The situation in which several students read a book together is suitable to use our system (Figure 3). Each student read a chapter and share the knowledge with others. They need to summarize the contents by underlining the important points or taking notes. The system will provide a note map to them when they complete their reading tasks. They present the contents they read according to the note map. All students share their note map and add a comment to others’ note map.

Students directly find the key points of the chapter which s/he did not read and improve their learning efficiency through obtaining the note-map of other students. Since we record the students’ reading track and display it on the note-map, the students will see the relationship among the pages on the note-map. Students can use this function to refer to others’ reading trajectories while reading notes. In other words, students can select the pages to read according to their needs, such as, s/he extracts the closely related pages to read according to the reading trajectories, if there is knowledge point that s/he wants to understand in detail.
From the standpoint of the student sharing her/his notes, since they will share their notes with other students, they should stand as a teaching point of view and be careful about writing the notes. This is will be a good reflection on their study and will gain a deeper understanding by teaching others what they have learned. Every student can edit the notes that are shared; the notes will get error correction from other students.

From the standpoint of a student who is reading other students’ notes, s/he can shorten her/his reading time. A book that takes seven days to read may be finished in three days by reading other students’ notes and listening to their explanations. After knowing the content of the whole book, students can read and understand it carefully according to their own needs.

4. Contributions and Future Work

Several ideas of collaborative note-taking systems are independent systems, and few are built on the e-books. A visual note-map that we proposed is based on separate student’s reading behavior on DITel e-book system. It not only provides sharing capability but also creates an opportunity for each student to reflect on her/himself. Furthermore, a scenario of usage that supports collaborative reading was presented in this paper. The collaborative reading which is a kind of method to improve reading efficiency and encourage student’s engagement has been wildly used in educational fields. In the future, the effectiveness of the system will be tested in the experiment.

References


Extending Deep Knowledge Tracing: Inferring Interpretable Knowledge and Predicting Post-System Performance

Richard SCRUGGS*, Ryan S. BAKERa & Bruce M. MCLARENb
aGraduate School of Education, University of Pennsylvania, United States
bHuman-Computer Interaction Institute, Carnegie Mellon University, United States
*rscr@gse.upenn.edu

Abstract: Recent student knowledge modeling algorithms such as Deep Knowledge Tracing (DKT) and Dynamic Key-Value Memory Networks (DKVMN) have been shown to produce accurate predictions of problem correctness within the same learning system. However, these algorithms do not attempt to directly infer student knowledge. In this paper we present an extension to these algorithms to also infer knowledge. We apply this extension to DKT and DKVMN, resulting in knowledge estimates that correlate better with a posttest than knowledge estimates from Bayesian Knowledge Tracing (BKT), an algorithm designed to infer knowledge, and another classic algorithm, Performance Factors Analysis (PFA). We also apply our extension to correctness predictions from BKT and PFA, finding that knowledge estimates produced with it correlate better with the posttest than BKT and PFA’s standard knowledge estimates. These findings are significant since the primary aim of education is to prepare students for later experiences outside of the immediate learning activity.

Keywords: Deep knowledge tracing, Bayesian knowledge tracing, dynamic key-value memory networks, performance factors analysis, latent knowledge

1. Introduction

In the last five years, a revolution has been underway in student knowledge modeling. For two decades, a dominant algorithm, Bayesian Knowledge Tracing (BKT; Corbett & Anderson, 1995) was the primary option. Other algorithms, both variants on BKT and – after 14 years – variants on logistic regression (e.g., Pavlik, Cen, & Koedinger’s [2009] Performance Factors Analysis [PFA]) and item response theory (Wauters, Desmet, & Van Den Noortgate, 2010) competed with BKT, but the differences in performance between algorithms were small (Gong, Beck, & Heffernan, 2010).

Then, after two decades, Piech et al. (2015) published an exciting new algorithm, Deep Knowledge Tracing (DKT), based on recurrent neural networks, along with initial evidence that its performance at predicting immediate correctness was substantially higher than BKT. Though the difference appears to be somewhat smaller than initially reported, there nonetheless appeared to be a benefit to using DKT instead of BKT (Xiong, Zhao, Van Inwegen, & Beck, 2016; Khajah, Lindsey, & Mozer, 2016). Several papers quickly emerged, proposing extensions and improvements to DKT (e.g., Cheung & Yang, 2017; Yeung & Yeung, 2018; Zhang, Shi, King, & Yeung, 2017; Zhang, Xiong, Zhao, Botelho, & Heffernan, 2017), while other papers explored the utility of additional machine learning methods in predicting problem correctness (e.g., Jiang, Ye, & Zhang, 2018; Lincke, Jansen, Milrad, & Berge, 2019).

However, DKT and its successor algorithms seemed to have two significant limitations relative to earlier approaches. First, DKT as originally implemented produced unstable performance, with oscillating predictions that sometimes went down after producing a correct answer. Yeung and Yeung (2018) proposed a regularization procedure which addresses this limitation. Second, DKT produced only predictions of correctness rather than an estimate of student knowledge on specific human-interpretable skills (see discussion in Pelánek, 2017). The first half of this limitation was addressed by J. Zhang and colleagues (2017), who introduced a skill-item matrix in their approach, Dynamic Key-Value Memory Networks (DKVMN). DKVMN produces predictions of latent
knowledge, but with reference to a machine-generated set of skills rather than a human-designed set of skills (the same interpretability limitation is seen in recent approaches that modify BKT to bring its performance into line with DKT – e.g., Khajah et al., 2016).

Indeed, despite DKVMN’s stated goal of inferring external knowledge, the initial paper on this algorithm did not attempt to actually predict performance on external measures of knowledge, sticking to the now-dominant paradigm of predicting immediate correctness. What’s more, to the best of our knowledge, none of the dozens of papers of DKT and its successors have explicitly attempted to measure how well these approaches perform at inferring the knowledge that is carried outside the learning system, through a post-test or other methods, in strong contrast to the early work on BKT, where considerable attention was paid to this goal (e.g., Corbett & Anderson, 1995; Corbett & Bhatnagar, 1997. Also see Pardos, Gowda, Baker, & Heffernan, 2011, for an example including PFA). While other recent papers attempt to tie learning data to skill proficiency (e.g., Wong et al., 2017; Yeung, 2019), their approaches focus on allowing algorithms to discover skills and skill relationships rather than linking back to known and interpretable external skills.

To address this issue, in this paper we attempt to reconnect student knowledge modeling with its initial roots in predicting student knowledge that goes beyond the confines of the learning system. First, we propose a very simple extension that can be applied to DKT, DKVMN, and other algorithms in this family, to enable the algorithms to predict external performance on externally-defined and meaningful skills. This extension consists solely of taking the real-time predicted probability of correctness over all items that a student answered that have been tagged with each external skill, and then calculating the mean of those values for each student, within each skill. To some extent, this follows the “correct first attempt rate” used by Yu et al. (2010) in their KDD Cup winning entry, combining students' performance on all the problems that they attempt. It also captures a student’s degree of difficulty in getting to mastery within the system as well as their final state; as Corbett and Bhatnagar (2017) note, final mastery estimates can be incomplete estimates of the knowledge a student carries out of a system when that system has enough practice for most students to reach high within-system proficiency. Although this paper applies this extension solely to predicting performance on an external test, this extension could also be used to report current skill levels to students and teachers in a meaningful fashion (in skill bars, perhaps, as seen in Cognitive Tutors and other platforms [Koedinger & Aleven, 2007]).

Second, we apply this extension to the outputs of DKT and DKVMN, and compare their performance on an external post-test measure of student knowledge to the classic BKT and PFA algorithms. Since it is classic BKT that has extensive evidence for making latent knowledge predictions that are both interpretable and predict post-tests effectively, we use BKT’s original formulation rather than modern extensions (i.e., Khajah et al., 2016). Third, we apply this extension to BKT and PFA as well, finding that the same extension improves prediction of post-test performance for these algorithms as well.

In the remainder of this paper, we present each algorithm in greater detail, present the data set that these algorithms will be compared within, discuss results, and then conclude with a discussion of implications and future extensions.

2. Algorithms Studied

2.1 Deep Knowledge Tracing

Deep knowledge tracing (DKT) uses recurrent neural networks to model student performance learning (Piech et al., 2015). It does not provide estimates of latent student knowledge (unlike BKT), and does not provide estimates of performance for a skill in general (unlike PFA), only predictions of correctness for each actual problem in the data. We implemented DKT using code from Yeung and Yeung (2018), who added extensions to the original method (Piech et al., 2015). The extensions address irregular fluctuations in correctness probabilities as students complete the learning activities and eliminate occasional instances where estimated correctness probabilities either decreased after correct answers or increased after incorrect answers.
In order to generate predictions of external knowledge, we took the probability of correctness over all items that a student answered from each skill, and then calculated the mean of those values for each student, within each skill. These resulting means were then used as knowledge estimates. We refer to these knowledge estimates as coming from mean-DKT.

2.2 Dynamic Key-Value Memory Networks for Knowledge Tracing

Dynamic Key-Value Memory Networks (DKVMN) represents states and the relationships within them with two matrices, one for storing internally-derived knowledge components and KC-item mappings and the other for storing the mastery associated with each knowledge component (J. Zhang et al., 2017). While DKVMN produces latent knowledge estimates like BKT, unlike BKT these estimates cannot be straightforwardly mapped back to externally-defined skills, as a new skill-item mapping is distilled bottom-up by DKVMN. Therefore, in order to map DKVMN’s estimates back to the posttest, we used the same approach as for DKT: we calculated the mean probability of correctness for each item associated with each skill for each student and used these means as knowledge estimates, referring to those estimates as coming from mean-DKVMN. Code from J. Zhang and colleagues (2017) was used to implement DKVMN.

2.3 Bayesian Knowledge Tracing

Bayesian Knowledge Tracing (BKT; Corbett & Anderson, 1995) is an algorithm that infers the probability that students have mastered a skill and the probability they will correctly answer a question which demonstrates that skill. BKT is often thought to differ from other knowledge and performance modeling algorithms in that it explicitly models latent knowledge as well as predicting future correct performance (e.g., Baker, 2019), differentiating between the two with estimates of slip and guess that reflect how performance may not entirely match knowledge. In this study, BKT was implemented using code from Baker et al. (2010), which estimates guess, slip, initial knowledge, and learning transition probabilities for each skill. The parameters were bounded to avoid model degeneracy (Baker, Corbett, & Aleven, 2008), with a floor of 0.01 for all probabilities, a ceiling of 0.3 for guess and slip, and all others having a ceiling of 0.99.

The parameter estimates were applied to the problem data using Excel and the final probability of having learned each skill was recorded for each student. In addition to taking the final probability estimated for each skill, we also calculated knowledge estimates by computing the mean correctness probability for each skill for each student across all of that student’s attempted problems, for comparability to the approach used for DKT and DKVMN. We refer to this variation as mean-BKT.

2.4 Performance Factors Analysis

Performance Factors Analysis (PFA), pioneered by Pavlik, Cen, and Koedinger (2009), models and predicts student performance using a logistic regression equation that models changes in performance in terms of the number of student successes and failures that have occurred for each skill. PFA estimates the probability of correctness, which is considered as an estimation of learning (Pavlik et al., 2009). In this study, the algorithm was implemented in Excel following the formulas in Pavlik et al. (2009), and using the Excel equation solver to determine optimal parameter estimates. The final learning probability was recorded for each skill for each student.

As with the other algorithms, we also calculated knowledge estimates by computing the mean correctness probability for each skill for each student across all of that student’s attempted problems. We refer to this variation as mean-PFA.

3. Participants and Data Collection

1 Using DKT, we were unable to calculate valid correctness predictions for 22 problem attempts, out of a total of 70,552 attempts. Those invalid attempts were omitted.
Data from the present study were originally collected for a series of studies, conducted across three semesters, on the effectiveness of erroneous examples on student learning (Richey et al., 2019). The studies aimed to improve students’ understanding of decimal numbers and their operations, particularly relating to several common misconceptions held by students (Stacey, Helme, & Steinle, 2001). Participants in the study were sixth-grade students at five urban and suburban schools in the northeast U.S. Data were collected over a six-day period in each study. The materials used in the three studies were the same except that the second and third semester versions of the study had twelve more practice problems than the first. The students received slightly different educational materials depending on whether they were assigned to an erroneous examples group or a more standard problem-solving group. Students in both groups received the same problem content, but erroneous examples problems began by describing a hypothetical student who had answered the problem incorrectly. In both groups, students were then asked to solve a problem (in the case of erroneous examples, this meant finding and correcting the error) and answer an explanatory multiple-choice question about their reasoning. If students responded correctly, they proceeded to the next problem; if they responded incorrectly, they were prompted to answer the incorrect sub-problem(s) with errors again until they got it correct. The materials used did not contain any hints.

A total of 598 students were included in the studies, with 287 students in the erroneous examples group and the remaining 311 in the problem-solving group.

All materials and posttests were delivered through the Tutorshop learning management system, which recorded students’ interactions (Aleven, McLaren, & Sewall, 2009). The materials were developed with the Cognitive Tutor Authoring Tools (Aleven et al., 2016). More information about the materials is available in Richey et al. (2019), McLaren, Adams, and Mayer (2015), and Adams et al. (2014). More information about the skills and their relationship to the misconceptions is available in Nguyen, Wang, Stamper, and McLaren (2019).

Students in the study were given 36 (208 students) or 48 (390 students) problems aimed to increase their understanding of decimal numbers. Each of the 36 or 48 problems comprised several subproblems. The problems covered four different skills:

- Ordering decimal numbers by magnitude
- Placing decimal numbers on a number line
- Completing a sequence of decimal numbers
- Adding two decimal numbers

In total, our data set contained 70,552 student attempts at subproblems: 28,908 for ordering decimals, 24,115 for placement on number line, 10,762 for completing the sequence, and 6,767 for decimal addition.

After students completed the problems, their understanding was checked with a 43-item posttest, which tested the four skills. Different numbers of items were used for different skills, in accordance with the number of common misconceptions which were presented for each skill (Richey et al., 2019): 22 items addressed ordering decimals \((M = 0.71, S.D. = 0.26)\), six addressed placement on a number line \((M = 0.53, S.D. = 0.31)\), four addressed completing the sequence \((M = 0.59, S.D. = 0.28)\), and eleven addressed decimal addition \((M = 0.66, S.D. = 0.23)\).

4. Algorithm Application

First, we simplified the students’ interaction data, keeping only whether students answered correctly or incorrectly on their first attempt at each problem, in line with common practice in student latent knowledge estimation (Corbett & Anderson, 1995; Pavlik et al., 2009; Piech et al., 2015; J. Zhang et al., 2017). Interaction attempts were then labeled with their associated skill. Next, we trained the set of different student latent knowledge estimation algorithms listed above, using all of the first-interaction data as training data. The implementations of DKT and DKVMN that we used expected separate training and test data sets, however, in this case we used the same data for both sets, since our goal is to understand performance on entirely new external data (posttests) rather than predict future within-system performance. After the algorithms were trained, we derived knowledge estimates for each student using each algorithm. The basic process of training and gathering knowledge estimates...
was generally similar from algorithm to algorithm, but differed based on how the algorithms treat (or fail to treat) latent knowledge.

5. Statistical Comparisons Between Algorithms

After using the four algorithms to produce estimates of latent knowledge for each student and each skill, the estimates were compared. First, we calculated Pearson correlations between each algorithm’s knowledge estimates and the posttest scores. As all measurements came from the same population of students, we were able to use a statistical test of the difference in statistical significance between correlations for correlated samples to compare the various correlations to each other (Ferguson, 1976). This test tells us whether one correlation (i.e. one model’s ability to predict the post-test) is statistically significantly higher than another correlation (i.e. another model’s ability to predict the post-test).

After comparing each combination of algorithms, we performed the Benjamini-Hochberg post hoc control procedure to control for the use of multiple comparisons (Benjamini & Hochberg, 1995; Benjamini & Yekutieli, 2001). This procedure reduces false positives by increasing stringency as more comparisons are performed, maintaining the same false discovery rate regardless of how many statistical tests are conducted.

6. Results

Table 1. Pearson correlations between knowledge estimates and posttest scores

<table>
<thead>
<tr>
<th></th>
<th>Ordering Decimals</th>
<th>Placement on Number Line</th>
<th>Complete the Sequence</th>
<th>Decimal Addition</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean-DKT</td>
<td>0.71</td>
<td>0.64</td>
<td>0.34</td>
<td>0.48</td>
</tr>
<tr>
<td>mean-DKVMN</td>
<td>0.72</td>
<td>0.62</td>
<td>0.35</td>
<td>0.56</td>
</tr>
<tr>
<td>PFA</td>
<td>0.28</td>
<td>0.33</td>
<td>0.10</td>
<td>0.26</td>
</tr>
<tr>
<td>mean-PFA</td>
<td>0.69</td>
<td>0.64</td>
<td>0.36</td>
<td>0.49</td>
</tr>
<tr>
<td>BKT</td>
<td>0.44</td>
<td>0.43</td>
<td>0.28</td>
<td>0.49</td>
</tr>
<tr>
<td>mean-BKT</td>
<td>0.65</td>
<td>0.52</td>
<td>0.28</td>
<td>0.44</td>
</tr>
</tbody>
</table>

Table 2. T-scores of correlations between comparisons. * indicates B-H significance at 0.05 level.

<table>
<thead>
<tr>
<th></th>
<th>mean-DKVMN</th>
<th>PFA</th>
<th>mean-PFA</th>
<th>BKT</th>
<th>mean-BKT</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean-DKT</td>
<td>1.65</td>
<td>-14.35*</td>
<td>-1.74</td>
<td>-10.59*</td>
<td>-3.74*</td>
</tr>
<tr>
<td>mean-DKVMN</td>
<td>-14.67*</td>
<td>-3.22*</td>
<td>-11.86*</td>
<td>3.84*</td>
<td>9.89*</td>
</tr>
<tr>
<td>PFA</td>
<td>14.17*</td>
<td>3.84*</td>
<td>9.89*</td>
<td>6.43*</td>
<td></td>
</tr>
<tr>
<td>mean-PFA</td>
<td>-10.77*</td>
<td>-7.94*</td>
<td>-7.14*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BKT</td>
<td>-7.94*</td>
<td>-7.14*</td>
<td>-6.95*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>mean-DKVMN</th>
<th>PFA</th>
<th>mean-PFA</th>
<th>BKT</th>
<th>mean-BKT</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean-DKT</td>
<td>-1.35</td>
<td>-9.72*</td>
<td>0.08</td>
<td>-8.07*</td>
<td>-6.95*</td>
</tr>
<tr>
<td>mean-DKVMN</td>
<td>-8.85*</td>
<td>1.26</td>
<td>-7.72*</td>
<td>-5.31*</td>
<td>-5.04*</td>
</tr>
<tr>
<td>PFA</td>
<td>9.79*</td>
<td>2.53*</td>
<td>5.04*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean-PFA</td>
<td>-7.94*</td>
<td>-7.14*</td>
<td>-7.14*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BKT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1 shows the correlation between each algorithm’s within-tutor knowledge estimates and posttest performance for each skill. Table 2 shows t-scores of the resulting comparisons, with an indication of which tests remained statistically significant after performing the Benjamini-Hochberg control, with FDR (false discovery rate) set to 0.05, equivalent to a p-value of 0.05 for a single test. Results for three skills were broadly similar, with mean-DKT, mean-DKVMN, and mean-PFA producing better estimates than traditional PFA and BKT. Mean-BKT produced estimates that outperformed traditional BKT and PFA in several cases, but generally performed lower than mean-DKT, mean-DKVMN, and mean-PFA.

For Ordering Decimals, mean-DKT (r=0.71) and mean-DKVMN (r=0.72) produced the closest knowledge estimates to the posttest scores. Mean-PFA (r=0.69) produced estimates that were significantly worse than mean-DKVMN, but not significantly different from DKT.

Mean-BKT’s (r=0.65) estimates, although close to mean-PFA, were significantly worse than that algorithm, as well as mean-DKVMN and mean-DKT. Estimates from BKT (r=0.44) and PFA (r=0.28) were significantly worse than the other algorithms, with BKT significantly better than PFA.

All algorithms produced worse results on Placement on Number Line, although the order of the groups did not notably diverge from Ordering Decimals. Mean-DKT (r=0.64), mean-PFA (r=0.64), and mean-DKVMN (r=0.62) all produced estimates that did not differ significantly from each other, but beat mean-BKT (r=0.52), BKT (r=0.43), and mean-PFA (r=0.33). Mean-BKT (r=0.52), however, performed better than PFA and BKT.

Complete the Sequence saw all algorithms struggle compared to the first two skills. Mean-PFA (r=0.36), mean-DKVMN (r=0.35), and mean-DKT (r=0.34) performed approximately equally. BKT (r=0.28) was able to produce estimates that were close to the top three and not significantly different from mean-DKT, although its prediction of the post-test was still statistically significantly worse than mean-PFA and mean-DKVMN. Mean-BKT’s (r=0.28) estimates did not significantly differ from BKT, but were worse than mean-PFA, mean-DKVMN, and mean-DKT. PFA (r=0.10) performed significantly worse than all other algorithms for this skill.

For Decimal Addition, mean-DKVMN (r= 0.56) achieved significantly better prediction of the post-test than the other algorithms. In turn, mean-PFA (r=0.49), BKT (r=0.49), and mean-DKT (r=0.48) achieved significantly better prediction than PFA (r=0.26). Although BKT’s estimates correlated better with the posttest than mean-BKT (r=0.44), that difference was not statistically significant, but mean-PFA produced significantly better estimates than mean-BKT. This finding may seem non-intuitive, since BKT and mean-PFA achieved the same correlation; it is due to there being a higher correlation between mean-PFA and mean-BKT than between BKT and mean-BKT.

7. Discussion and Conclusions
Although Deep Knowledge Tracing and Dynamic Key-Value Memory Networks were not designed to produce estimates of latent knowledge for predefined skills, our approach was able to convert performance predictions made by these algorithms into knowledge estimates, which achieved reasonable correlation to student scores on an external posttest. These estimates were more accurate at predicting the external posttest than estimates from Bayesian Knowledge Tracing, which was designed with the aim of estimating the state of students’ knowledge. Mean-DKVMN and mean-DKT’s estimates were comparable to or perhaps a little better than estimates provided by the classic knowledge modeling algorithm Performance Factors Analysis. In other words, though deep learning-based models might have been thought to mainly capture performance within the system, with a simple adjustment they are also better at inferring the knowledge students carry out of the learning system.

Curiously, PFA only performed comparably to mean-DKT and mean-DKVMN when the same adjustment was made to PFA as was necessary for DKT and DKVMN: averaging estimates across the actual problems, rather than simply taking the final estimate of knowledge for the skill. Explaining this finding may require going back to findings from some of the earliest work in this area. Corbett and Bhatnagar (1997) noted that if mastery learning is used – where a student continues to work within a learning system until the BKT estimate of their knowledge is very high (in that case 0.975) – there is very little variance in the final estimates of student knowledge (as all estimates are above 0.975). However, performance is not always equally high in external post-tests; BKT estimates for students driven to mastery tend to over-estimate post-test performance (Corbett & Anderson, 1995; Corbett & Bhatnagar, 1997). Notably, over-prediction appears to be more characteristic of cases where students had more remedial practice (Corbett & Anderson, 1995). Although the data set used in the current paper did not involve mastery learning, there was a sufficiently large amount of practice in that system (9 to 12 problems per skill for each student) to have caused similar phenomena. For three of the four skills, nearly all final knowledge estimates asymptotically approached either 0 or 1, although students rarely got all posttest items correct or all incorrect. By averaging estimates across problems, we capture student knowledge throughout the learning process rather than apparent knowledge at the end – capturing lower performance on the eventual path to mastery – which appears to be a better estimate of the knowledge students carry out of their learning experience. However, this does not completely explain our results: for Ordering Decimals, no students had final knowledge estimates greater than 0.95 or less than 0.05, but our adjustment still significantly improved the posttest correlations for that skill.

The same adjustment of averaging estimates across actual problems rather than using final knowledge estimates led to better performance for BKT as well as PFA, although not to the same degree. In this paper, the original version of BKT was used. Recent work has suggested that BKT performs better at predicting within-system correctness if several adjustments are made (i.e., Khajah et al., 2016), though still not as well as DKT. It is possible that a version of BKT adjusted in this fashion may perform more comparably to mean-DKT, mean-DKVMN, and PFA for predicting the post-test. However, the very adjustments necessary in Khajah et al. (2016) eliminate some of the benefits – such as interpretable estimates of student knowledge on expert-defined skills – that have made BKT an attractive alternative for practical use.

One of the major arguments in favor of Bayesian Knowledge Tracing has been its interpretable latent estimates – separate from performance. This paper’s findings suggest that BKT’s latent estimates may not be as useful as thought. BKT does more poorly at estimating an external post-test measure than a reasonable transformation of modern deep learning based algorithms, as well as a more traditional competitor, PFA. Combined with evidence that BKT does more poorly at forecasting time until mastery than PFA (e.g., Slater & Baker, in press), and evidence that classical BKT does more poorly at forecasting future performance within a learning system than DKT or DKVMN (Khajah et al., 2016; J. Zhang et al., 2017), it appears that BKT’s use as a primary knowledge modeling algorithm may be coming to an end. With the simple modification to DKT or DKVMN provided here, assessments of specific understandable skills can be provided to teachers and students, one of the core uses of BKT (Koedinger & Aleven, 2007), and these estimates are more predictive of post-test performance than BKT’s estimates.

Our findings should not be interpreted as indicating that Bayesian Knowledge Tracing has no use, however. Bayesian Knowledge Tracing still offers the advantage of interpretable parameters, and there are cases – particularly when one wants to understand which skills have low learning rates or high slip rates (e.g., Agarwal, Babel, & Baker, 2018), where BKT may be very useful. In addition, distillations of Bayesian Knowledge Tracing, such as student-level contextual slip, remain useful
predictors of long-term outcomes (e.g., San Pedro, Baker, Bowers, & Heffernan, 2013). At this point, however, its shortcomings in predictive accuracy make it harder to justify their use in cases where model structure does not need to be explained.

Of course, no single result is definitive, and more research is needed to establish our findings here as conclusive. This study only investigated data from students’ experiences learning decimals in one tutoring system, comparing learning estimates with a single posttest. Our findings, particularly regarding BKT’s ability to predict external measures, should be replicated with different student populations and in different domains. However, the results should be encouraging to researchers interested in using DKT, DKVMN, and other cutting-edge knowledge tracing algorithms to infer knowledge, rather than just predicting performance within-system.

There has been considerable work over the last several years to discover which student knowledge model is best at predicting future correctness within intelligent tutoring systems. In Corbett and Anderson’s (1995) original vision for student knowledge modeling, as much attention was given to prediction of performance outside the learning system as within it. This seems appropriate, given that the true goal of education is not what students can do during learning, but what they can do beyond and going forward. In this paper, we find that simple enhancements make it possible for recent emerging performance prediction algorithms to also effectively predict knowledge that extends outside the tutoring system. The simple solution provided here will almost certainly fall short of the best that can be done. We hope that in the years to come as much attention will be provided to the problem of predicting long-term and system-external performance as predicting immediate correctness has received recently. Ultimately, the goal of student knowledge modeling should be to infer knowledge, not just predict performance. Happily, it seems like the newest student knowledge algorithms can successfully do this, with only a modest adjustment.

References


Slater, S., Baker, R.S. In press. Forecasting Future Student Mastery. To appear in *Distance Education*.


Introducing a Mock Technique into a Learning Support System for Program Design Based on Testability

Masaya MURAMATSUa*, Yasuhiro NOGUCHIb, Satoru KOGUREa, Koichi YAMASHITAb, Tatsuhiro KONISHIa, & Yukihiro ITOHC

a Faculty of Informatics, Shizuoka University, Japan
b Faculty of Business Administration, Tokoha University, Japan
c Shizuoka University, Japan
* muramatsu.masaya.16@shizuoka.ac.jp

Abstract: While software development requires automated testing skills, there are few opportunities for test education during a novice programming learning phase. Therefore, we constructed a learning support system that can conduct an exercise to analyze and improve program design from the viewpoint of Testability in the design of the component to be tested. However, our previous system did not support design improvement using mock techniques, and the method of improvement that learners could choose was limited. Therefore, in this paper, we extended the system to support learning for analyzing and improving the design of the test target component with mock techniques. We experimented to evaluate the learning effect of the system whether the learners increase their skills of analyzing and improving the design of test target components, even in the situation where applying object-oriented design principles are required for the learners.

Keywords: Programming education, learning support environment, unit test, testability, program design, mock

1. Introduction

In recent years, as the software has become larger and more complex, the importance of program verification has increased. While knowledge and practical skills for testing are required, there are few opportunities for testing education at early programming learning stages, such as the college study of computer science. Therefore, research have been conducted to support the learning and implementation of unit test automation, such as Arcuri, Fraser, and Galeotti (2014) and Proulx and Jossey (2009).

Noguchi, Ihara, Kogure, Yamashita, Konishi, and Itoh (2019) developed the Testability Learning System (TLS), which supports the learning of the designs of test target components, focusing on the index of Testability organized by Bach (1999). By using TLS, learners can acquire how to design testable components in two phases: analysis phase and improvement phase. In the first phase, the learners analyze the causes of difficulties to test the test target component. In the second phase, the learners improve the designs of the component based on the causes analyzed. In both phases, the learners complete the tasks with the guiding of predefined feedbacks from the system. However, TLS does not support mock techniques thereby limiting design improvement methods.

There are two problems with the lack of methods of improvement with mock techniques. First, learners must decide from limited candidates for the methods of improvement for the analyzed cause and it causes that learners narrow the understanding of the relationship between the cause and the appropriate component design that solves the cause. Second, the test target components to be practiced are limited because some components are difficult to be tested without mock techniques. For instance, a component that receives essential data through calling an external API is a lack of Controllability, when specific data are required to invoke the component’s behavior to be tested. If learners cannot use mock techniques, they can choose only a method of improvement that moves the codes of calling the external API to the outside of the component and prepares some interfaces of the component to be set every essential data (e.g. setters, method parameters, etc.). The limitation of methods of improvement has two problems mentioned above. Learners who learned the method as only a choice to improve the lack of controllability cannot notice that the method makes the component depend on other components by
overusing control coupling. Also, a component that uses different data based on the condition of the component cannot be considered as a sample component at the learning. In this case, the component cannot sustain the provided function because the method moves most code fragments to out of the component. Moreover, most components excluding simple codes only for programming exercises categorized as such components.

Redefinition, polymorphism, and dynamic binding are important features of object-oriented design for testable components by using mock techniques (hereinafter, "OOD features"), and many researchers have conducted on the difficulty of the learning and some educational methods (e.g., Or-Bach, & Lavy, 2004; Hadar, & Leron, 2008; Arif, 2000; Bergin, 2003; Schmolitzky, 2006; Benaya, & Andzur, 2008; Alkazemi, & Grami, 2012; Daniel, 2019). Therefore, it is important to provide some supports to understand and utilize the OOD features while they analyze and improve the design of the test target components.

In this paper, we extended previous TLS so that learners can analyze and improve the design of the test target component with mock techniques. The extended TLS prepares the learning flow for supporting such analysis and improvement, and the system can support the learners in completing the flow. We conducted experiments to evaluate the extended system and verify the learning effects of the extended portion from the experimental results.

2. Related Works

For the learning and implementation of unit test automation, many research focused on the testing components to reduce the design/implementation costs (e.g., Schroeder, & Rothe, 2005; Allowatt, & dwards, 2005; Teixeira, & Silva, 2018). As for the practice for early programmers with the design of the test target components, there are some cases that Test Driven Development (TDD) activities used in a programming exercise (e.g., Janzen, & Saiedian, 2008; Funabiki, Matsushima, Nakanishi, Watanabe, & Amano, 2013). In these practices, however, their students are often provided sample projects including full set of automated unit tests. It means that testing/tested component design is provided for students and they focused on implementing the behaviors of the tested components and adding extra test cases. There are some research to support TDD activities for early programmers, such as Johnson and Kou (2007). As mentioned above, these research do not focus on supporting learning of design and implementation for the test target component.

In this paper, we extended TLS (Noguchi, et al., 2019) for supporting learners to improve their design of components with mock techniques. We summarize the purpose and functions of the previous TLS and clarify our contributions in this paper. The previous TLS focused on three indicators of Testability—Controllability, Observability, and Decomposability—that are related to the interface design of the component to be tested. Figure 1 shows the relationship between the three indicators and automated testing. Controllability is an indicator of whether the test driver can control the test target component to cause its behavior to be tested. Observability is an indicator of whether the test driver can observe the test target component’s behavior to be tested. Decomposability is an indicator of whether the test target component consists of only the codes to be tested.

Figure 1. Testability Indicators
In the previous TLS, learners can choose an exercise whose sample program code includes a test target component with some difficulties to be tested, and the learners analyze the causes of the difficulties and improve the design of the test target component to resolve the difficulties to be tested. The learners can improve the design based on their analyzed causes of the difficulties by choosing the methods of improvement provided by the system. At each step of the analysis and design improvement, the system evaluates the learners’ choices and provides feedback. When the learners completed the steps, their improved program code can be downloaded as an Eclipse project, and the learners can confirm whether they can write automated tests for the improved test target component in Eclipse by executing the automated tests. However, the previous TLS cannot provide methods of improvement with mock techniques, because the TLS focused on the learners at early programming learning stages who have not learned the OOD features that are essential for using the methods.

Regarding the difficulties in learning the OOD features, Liberman, Beeri, and Kolikant (2011) analyzed an object-oriented programming course where 22 in-service CS teachers whose programming mother tongue was procedural (Pascal, C, or Basic) learned object-oriented programming. Then, they categorized the difficulties in learning inheritance into four clusters: one, by not presenting a clear model, participants develop their model for a new situation from past knowledge, etc.; two, by explaining it with similar concepts that learners are familiar with, they stumble during implementation; three, by confusing inheritance of the is-a relation with the composition of the has-a relation; and four, by assuming that an object of a class exists when the definition of the class is written, even before executing construction code for the object. In our study, we provide learners an obvious design model that satisfies Testability to support automated testing. Moreover, both the correctness of its design and the implementations can be evaluated according to whether the automated testing can be applied or not. Thus, in our study, learners can avoid the first and the second difficulties because they are guided by the obvious design model and teachers are not required to explain design principles by using similar concepts that learners are familiar with. As for the third and fourth difficulties, learners are supported by providing feedback from the system about the design and implementation errors.

Nandigam, Tao, Gudiwada, and Hamou-Lhadj (2010) proposed that unit testing using a mock object framework is an effective practical subject for teaching object-oriented design principles. They organized the learning of object-oriented design principles using the mock object framework into four elements of instruction: the design must have an associated with unit test class, one must first prepare interface specification for a unit, the designer of each unit must strive to use interfaces instead of concrete classes, and everything that an object needs are passed into the object either as a parameter to the constructor or using an appropriate setter method. Whereas they aimed at learning object-oriented design principles, our study aimed at learning design concepts and practical techniques for improving component design based on Testability. Also, object-oriented design principles are only one of the necessary elements as the design improvement method. In our study, the mock object framework is a selectable element as one of the methods of design improvement when the learners themselves evaluate the component design based on Testability. If the design improvement by the mock object framework is effective, they should choose it; otherwise, they should choose a different design proposal. Moreover, since our proposed system covers the four main learning items Nandigam, et al. proposed, it may be able to be positioned as an educational support system of training developers in the object-oriented design principles the researchers proposed.

3. Learning Support Strategies for Design Improvement with Mock

3.1 Difficulties for Design Improvement with Mock

The OOD features enable developers to change the behavior of the invoked method without modifying the method name in code. It means that a developer can design a component that can be worked on a production purpose and a testing purpose without changing the codes of the component. As for the typical design for such components, Figure 2 shows a common design example used in testing frameworks such as mockito. Suppose that it is designed so that the behavior of the method to be tested on the test target component (“TestTargetMethod” in Figure 2) can be changed from the testing component (“Testing Component”) without modifying the code of the test target component. As for exchanging the production purpose and the testing purpose of the test target component, the test target
component can use the production component and the mock via the common interface ("Function to Replace" in "Interface"). The method to be tested invokes these components through the common interface, not using the concrete name of these components ("Production Component" and "Mock"). Additionally, the method to be tested can accept a parameter of the common interface ("obj: Interface" in "Test Target Component") which enables the testing component to insert the concrete component for the purpose ("Production Component" or "Mock").

![Figure 2. Design example with mock](image)

Regarding the difficulties for novice programming learners, they must learn the OOD features as their new knowledge, design the components based on these OOD features to realize automated testing, and modify the codes to realize the design of components from the original component which has difficulties of **Controllability** and/or **Observability**. In contrast with the learners’ situations of the previous TLS, there are major design changes from the original sample components, and collaborations of some part of design changes realize a part of testable features of the component. Thus, the learners sometimes lost steps in their improvement process; they cannot identify the design of components in their specific step; they cannot understand what part of design changes realizes a part of testable features of the component. As the result, it often happened that learners cannot notice their mistakes in the design changes themselves.

### 3.2 Design Improvement Process with Mock

We summarized the design improvement process with a mock with the following 6 steps:

**Step 1.** Choosing code fragments in the test target method should be replaced by a mock object that assigns parameters invoking the behavior to be tested or observes the results of the behavior to be tested.

**Step 2.** Defining the interface of the method invoked from the test target method that is commonly used in a production component and a mock.

**Step 3.** Moving the corresponding code fragments chosen at Step 1 into the production component, and change the invoking the production component via the interface defined in Step 2 from the test target component.

**Step 4.** Adding a parameter of the common interface into the test target method (or into the constructor of the test target component) to enable the testing component to insert a concrete object ("Production Component" or "Mock").

**Step 5.** Defining a mock object implemented based on the common interface, and, implementing the mock object based on the difficulties in the original test target component as it assigns parameters invoking the behavior to be tested or it observes the results of the behavior to be tested.

**Step 6.** Implementing the testing component that invokes the behavior of the test target method and verifies the responses of the method via inserted mock object.

At the end of Step 4, the design of components has been improved to enable the testing component to insert a concrete component. It means the improved code is possible to work as a
production purpose or a testing purpose. At the end of Step 5, a mock object has been prepared that can assign parameters and observe the results. At the end of Step 6, the testing component has inserted a mock object suitable for the testing situations that enable the testing component to control/observe the behavior of the test target object via the mock object.

There are various conditions for the design of the test target component, it is not necessary to complete the steps from Step 1. Moreover, learners must evaluate the design of the test target component and decide which step is appropriate for the component. Also, sometimes the learners decide to skip the steps that are not necessary to the design conditions.

3.3 Extending the Learning Flow

Figure 3 shows our extended learning flow from the previous TLS that did not support the design with mock techniques. In this study, we extended the step (1), (4)-B1, and (4)-B2 for improving the design with mock techniques. At the step (1), we explained the essential concepts for the design analysis and improvements in lecture-style method: Testability concepts, OOD features, design example with mock, and other knowledge for testable component design with mock techniques. In this study, we employ the existing steps (2), (3), (4)-A, and (5) for analysis and improvement of the design without mock techniques. The extended step (4)-B1 contains Step 1 and the step (4)-B2 contains Step 2-6 discussed in section 3.2.

Figure 3. Extended Learning Flow

3.4 Extended System’s Supporting Functions

3.4.1 Design Comparison View

To let learners identify their design of the component at the current step and the differences between the expected design of components and one at the current step, we extended the previous TLS to provide learners a view showing the both for comparing in Figure 4. This view shows the design of components in UML class diagram like format. In the view of their design at the current step, the mnemonics of the components follow the current real components the learners modified in their practice. Additionally, the system emphasizes the differences between the expected design and the design at the current step. Furthermore, the system also identifies where the learners should improve at the next step and the learners have successfully improved.
3.4.2 Message Feedback Strategies in Extended Steps

At each step, the system should provide feedbacks to learners who decide to choose a method of improvement. The feedbacks should encourage the learners by confirming their decision, and let the learners understand what features of the component are changed by their decided design change. In this study, the expected design is fixed and the system provides a limited number of candidates to the learners. Thus, the system’s feedback messages for possible learners’ choices can be predefined, and the code mnemonics corresponding to the expected design can be inserted into the feedback messages.

Regarding the steps extended in our study, at the step (4)-B1, the system provides candidates of code fragments in the test target method should be replaced by a mock object. If the learners choose incorrect code fragments (the learners can edit code fragments from the provided candidates in “Test Target Component Change Area” in Figure 5), the system provides learners feedback messages that explain how the difference between the code fragments as a mock object and the expected design. At the step (4)-B2 corresponds to Steps 2-6 in section 3.2, the learners can choose a candidate for the appropriate method of design improvement while checking the current design of components and the difference from the expected design by using the comparison view as Figure 5. If the learners choose the correct method of design improvement, the system provides learners feedback messages that what the features of the components are realized by the learners’ chosen design change. As for the incorrect method of design improvement, the system provides feedback messages for the problems let by the decision: syntax errors, exceptions occurred, losing the function of the original component design, losing the features of the component, and remaining difficulties for testing. Like Figure 5, the feedback messages generally consist of the four parts of information: instruction of the next step for the learner, explanations for confirming the learners’ decided design changes, explanations for the effects of the learners’ design changes, and explanations for the remaining problems in the component design caused by the inappropriate design changes.
4. Evaluation Experiment

4.1 Hypothesis

There are two hypotheses about the learning effects in exercises with sample components required design improvements by using mock techniques.

H1. Learners using our extended TLS increase their ability to analyze the causes of difficulties on the test target component rather than learners using IDE with a mock framework.

H2. Learners using our extended TLS increase their ability to improve the design of components rather than learners using IDE with a mock framework.

4.2 Evaluation Process

Figure 6 shows an overview of the evaluation process. At first, we collected 17 college students for the evaluation. 6 students are 3rd grade, 9 students are 4th grade, and 2 students are 1st grade in the graduate school. These subjects have basic knowledge of Java programming language confirmed 3 questions about them in the first 5 minutes in the pre-test. However, we eliminated 1 subject in the pre-test after our lecture for the subjects, because he acquired enough abilities only from the lecture.

In the pre-test, we also gave the subjects 2 problems improving the design of the component with mock techniques; one lacks Controllability, another lacks Observability. We categorized 8 subjects into the experimental group and 8 subjects into the control group based on the pre-test. The experimental group completes a new exercise improving the design of components with mock techniques with our extended TLS. The control group completes the same exercise with Eclipse IDE and the JUnit testing framework. While the experimental group received the tutorial for the TLS, the control group received the guide for analyzing/improving steps based on Figure 3. After these exercises, we gave a post-test with the same design improvement problems in their pre-test to both groups.

Regarding the evaluation of the design improvement problems in the pre-test and the post-test, the subjects answered on the answer sheets in 30 minutes. The answer sheet provided the subjects the codes of the original components. On the answer sheet, they can choose code fragments by enclosing with a square, and modify the codes into their improved codes. We evaluated the subject’s answers whether the subject can answer corresponding to the results of the analysis step (2) and (3) in Figure 3 and whether the subject can complete their design improvement corresponding to the step (4)-B1 and (4)-B2 including 6 steps discussed in section 3.2. Finally, we converted the evaluation scores of each portion on 10-scale points. As the result, a subject was evaluated in total of 40 points in both the post-test and the post-test. Additionally, we measured the difference between the scores of the pre-test and the post-test.

![Figure 6. Overview of the Evaluation Process](image)

4.3 Results

Table 1 shows the average scores in the pre-test and the post-test, the difference scores between both groups, and the difference scores between the pre-test and the post-test. The asterisks *, ** indicate that the coefficients are statistically different from zero at the 5 and 3 percent level based on Welch’s
two-sided t-test, respectively. The difference in the average scores in the pre-test was not statistically
different between both groups. Also, the differences from the score of the pre-test and the post-test
indicate the learning effects of the exercises in each group. Therefore, the difference in the learning
effects in the experimental group and the control group is evaluated by comparing with the both
difference scores.

For observing the learning effects for individual abilities, Table 2, Table 3 show the
differences in the average score of the analysis steps and the design improvement steps. Also, these
tables show the scores separately for the type of problems: controllability and observability.

For observing which step the subjects completed or failed, Figure 7 shows the number of
subjects who scored 0.5 or more points at each step in the post-test. Each step evaluated from 0 to 1 at an
interval 0.25. Besides, it shows the number of subjects who completed the whole steps with their score
more than the threshold. The threshold worked for focusing the design improvement portion on whether
the learner can proceed with the next step excluding the effects from negligible mistakes like
misspelling, slight syntax problems.

Table 1.

Differences in the average scores between the pre-test and the post-test (out of 40 points)

<table>
<thead>
<tr>
<th></th>
<th>Pre-test</th>
<th>Post-test</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>16.56</td>
<td>30.27</td>
<td>13.71</td>
</tr>
<tr>
<td>Control</td>
<td>15.62</td>
<td>19.82</td>
<td>4.20</td>
</tr>
<tr>
<td>Difference</td>
<td>0.94</td>
<td>10.45</td>
<td>9.51**</td>
</tr>
</tbody>
</table>

Table 2.

Difference in the average score of the analysis steps (out of 10 points for each problem)

<table>
<thead>
<tr>
<th></th>
<th>Observability</th>
<th>Controllability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-test</td>
<td>Post-test</td>
</tr>
<tr>
<td>Experimental</td>
<td>8.13</td>
<td>9.38</td>
</tr>
<tr>
<td>Control</td>
<td>5.94</td>
<td>6.56</td>
</tr>
<tr>
<td>Difference</td>
<td>2.19</td>
<td>2.82</td>
</tr>
</tbody>
</table>

Table 3.

Difference in the average score of the design improvement steps (out of 10 points for each problem)

<table>
<thead>
<tr>
<th></th>
<th>Observability</th>
<th>Controllability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-test</td>
<td>Post-test</td>
</tr>
<tr>
<td>Experimental</td>
<td>1.25</td>
<td>5.33</td>
</tr>
<tr>
<td>Control</td>
<td>1.44</td>
<td>1.46</td>
</tr>
<tr>
<td>Difference</td>
<td>-0.19</td>
<td>3.87</td>
</tr>
</tbody>
</table>

Figure 7. Number of Subjects who completed each step in the post-test
4.4 Discussion

As for the difference of the learning effect between the experimental group and the control group in Table 1, the score 9.51 that was statistically different from zero at the 3 percent level confirmed more learning effects of our extended TLS than the learning with only standard IDEs.

Regarding H1, Table 2 shows that the scores of the post-test for the experimental group were 9.38 and 10.0 for the problem that lacks Controllability and the problem that lacks Observability. These scores indicated the subjects in the experimental group almost completely resolved the problems in the post-test. Even though the differences in the learning effects: 0.63 and 2.50 in Table 2 did not have a significant difference by t-test, learning with our extended TLS has enough educational effects for increasing learners’ ability to analyze the causes of difficulties of the test target component to be automated tested with mock techniques.

Regarding H2, Table 3 shows that the scores of the post-test for the experimental group were 5.33 and 5.57 for the problem that lacks Controllability and the problem that lacks Observability. These scores indicated the subjects in the experimental group still made some mistakes in the post-test. As for the differences of the learning effects, the difference score 4.06 for the problem that lacks Observability had a significant difference by t-test. Even though the difference score 2.33 for the problem that lacks Controllability did not have a significant difference by t-test, both difference scores are not negative. Thus, H2 is partially supported by the results.

Besides, Figure 7 shows that a greater number of subjects in the experimental group satisfied at each step in the post-test rather than the subjects in the control group. It indicated that learning with our extended TLS has learning effects at every step rather than the learning with only standard IDEs. Also, the learning with our extended TLS may not support all learners but is potentially able to let learners increase their abilities to complete all steps because 2 subjects in the experimental group completed all steps. Regarding OOD features and Testability, learners choose appropriate code fragments based on the Decomposability perspective at Step 1; At Step 2, 3 and 5, learners require knowledge of inheritance and polymorphism to identify the common interface for multiple components and modify the code for invoking these components via the interface; At Step 4, learners improve the design of the test target component for realizing the concept of dynamic binding. The relatively small number of the subjects in both groups satisfied Step 5. It indicated that our extended TLS can support learners in Step 1, 2, 3, and 5 related to the concept of decomposability and inheritance. However, it also indicated that most learners lack of understanding of dynamic binding and the method of design improvement, and our extended TLS is also far from enough to support learners for the point.

5. Conclusion

In this paper, we extended previous TLS (Testability Learning System) for supporting learners to improve the design of the test target component with mock techniques. We summarized the design improvement process into 6 steps and implemented supporting features in the extended TLS. For modifying the component into the expected design, it is required that the learners can utilize some principles of the object-oriented design: inheritance, polymorphism, and dynamic bindings. Also, to choose the appropriate method of improvement, they must analyze the component design by the Testability perspectives: Controllability, Observability, and Decomposability. Although it is difficult for novice programming learners to apply these knowledges into the concrete design of components, our preliminary evaluation indicated some learning effects of the system to increase the learners’ practical skills of analyzing and improving the design of test target components. On the other hand, the evaluation indicated that most learners made mistakes the design improvement step related with dynamic binding even though the learners can receive some feedbacks by the system.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP18K11566.
References


mockito, Retrieved from https://site.mockito.org


Reimagining the conceptualisation, design and delivery of Learning Analytics

Aneesha BAKHARIA* & Linda CORRINb

aThe University of Queensland, Brisbane, Australia
bSwinburne University of Technology, Melbourne, Australia
*aneesha.bakharia@gmail.com

Abstract: Learning analytics applications have the potential to bring value to educational environments by providing the ability to analyse and present data about students in ways that are meaningful and actionable for educators. Recent studies have highlighted the importance of involving educators in the design process for these systems to ensure that what is developed meets their needs. However, this can add significant time and resource investment to a development project and requires the involvement of a range of people with data science and developer skills. Across various industries, new approaches are being imagined and trialled to reduce reliance on technical staff and long development processes when designing data-based systems instead enabling users to create a story from data without needing programming skills. In this paper we review the innovations introduced by DIVE, a mixed initiative visualisation and analysis application developed by a team at the MIT Media Lab, and discuss how these can be adapted to the field of learning analytics. DIVE includes a range of features that can provide value to learning analytics applications such as automated dataset statistical analysis, visualisation recommendation, and data story generation. An analysis of these features informs a set of guidelines which will enable the reconceptualisation of co-design sessions, reducing the time required to design learning analytics applications and facilitating the automated generation of prototypes, and in some cases full dashboards and analytic applications. The paper concludes with a discussion of future enhancements that could be made to the functionality of DIVE, such as the need for a semantic layer on trace and assessment data, to enable the translation of data into meaningful insights on learning.

Keywords: Learning Analytics, Dashboards, Visualisation Recommendation, Data Stories

1. Introduction

It takes time to design and mainstream a successful learning analytics application! The time investment is further extended when adopting an iterative co-design process, an approach which is rapidly growing in popularity in the learning analytics literature (e.g., Dollinger et. al., 2019, Shobani et. al., 2019). Co-design with educators ensures that the application directly addresses their requirements, resulting in an end product that provides more meaningful and actionable outputs to enable educators to support student learning. The increase in the use of co-design methods has also led to the development of tools to support the resulting design discussions. One example is a card-based design approach (Alvarez, et. al., 2020) where educators work through a series of design steps using cards relating to each area of the proposed application to identify useful components. While such approaches are able to trigger relevant discussion, they are only able to facilitate the abstract assembly of ideas. Much work still remains to develop the application and requires a diverse skill set (e.g., data science and developer skills) after the co-design session has ended to prototype, review and implement the application.

A further challenge in the design of learning analytics applications is enabling educators to make meaning from the analysis of different types of data (e.g. clickstream event data, assessment data, etc.) in a way that relates the outcomes of analyses of the data to what we know about how students learn. Within the learning analytics community this is known as the ‘clicks to construct’ issue (Knight & Shum, 2017). In order to be able to translate the data and analyses into meaningful insight there is an
important role that learning design and theory plays in providing context and an epistemological frame for interpretation (Knight, Shum & Littleton, 2014). Bringing together the data and the theory enables learning analytics practitioners and educators to build stories from the data that aid in understanding students’ actions and identifying appropriate interventions that can be made (Echeverria et al., 2018). However, this is a complex and sophisticated process which is especially hard to implement in systems that are designed to address a range of educator queries and needs.

While the complexity of learning is an ongoing challenge within the field of learning analytics, the other challenges identified here around the application development process are also present across a number of industries. As a result, the data science community continues to look for innovative ways to address these challenges, in particular to lower the technical and time barriers to allow users to easily build stories from data. As a field that grew from a multi-disciplinary foundation, learning analytics practitioners and researchers often look for innovations across disciplines and industries to inspire new ways of analysing and understanding education. In this paper we review recent innovations that have been made within data science research to improve the learning analytics application development workflow.

The review in this paper relates to three fundamental questions that are essential to re-conceptualising the design and delivery of learning analytics applications:

1. Is it possible to produce working prototypes within a design session between a learning analytics practitioner and educators?
2. Can learning analytic applications be automatically generated?
3. Can the generated learning analytics application address the ‘clicks to construct’ issue?

In order to answer these questions we first review a recently proposed application called DIVE which includes a workflow that would be suitable for the design of learning analytics applications. We then suggest improvements, and propose a set of detailed guidelines for adapting DIVE to a learning analytics setting and finally outline two proposed workflows within a higher education setting.

2. An Overview of DIVE

DIVE is a web-based platform that supports data ingestion, visualisation, analysis and storytelling. It was developed by Kevin Zeng Hu and César A. Hidalgo at the MIT Media Lab (https://www.media.mit.edu/projects/dive/overview/) to allow users to explore and use data to create stories without having to know how to code. DIVE provides a mixed-initiative workflow for analysts, including modern techniques such as visualisation recommendation and field level data model inference. In the DIVE evaluation study analysts were given a set of predefined visualisations and analytics to perform in DIVE and Microsoft Excel, with DIVE users showing a significant reduction in time to complete the tasks (Hu, et. al., 2018). DIVE has been selected to be reviewed as the data exploration workflow can be adapted to learning analytics and also because both the client and server-side code are open source. DIVE is implemented in Flask and React.

Figure 1 shows the DIVE user interface. The left bar (A) provides an intuitive way to navigate through the workflow provided by the application. The top bar (B) displays the current project and dataset being analysed. The main screen (C) displays the content of the selected workflow element with configurable parameters available on the right hand bar (D). The various sections of the DIVE user interface remain consistent across all supported workflows. Figure 2, illustrates the various workflows supported by DIVE.
2.1 Data Ingestion and Inspection

DIVE is able to import tabular data, perform data model inference and provide statistical analysis of each column. It categorises columns using a heuristic-based algorithm and uses this as additional information within any automated analysis. The basic categories are nominal, ordinal, and continuous along with a semantic categorisation which includes categorical, temporal, and quantitative. After data is imported, the basic statistical properties of each column are shown to the user. For categorical fields the number of unique values and the frequency of each value is displayed. Summary statistics are shown with simple histogram visualisations for ordinal and continuous fields.

2.2 Visualisation and Drilldown

DIVE provides mixed-initiative visualisation recommendations. The user specifies fields (or columns) of interest and the DIVE visualisation recommendation engine produces a set of visualisations that may be of interest to the user ranked by statistical measures. If no fields are selected, univariate descriptive visualisations are produced for all fields in the uploaded dataset. DIVE automatically generates charts which can be stored and reused within a story. There is no requirement for the user to know how to bind or encode data for display on a chart or have knowledge of using a charting library.
2.3 Analysis

DIVE is able to perform common statistical functions on the uploaded tabular data. The main functionality included is the ability to aggregate data (i.e., calculate mean and standard deviation), view the correlation between two variables, compare sub-groups using one-way or two-way ANOVA, and perform linear or logistic regression.

2.4 Story Creation and Sharing

Any visualisation or analysis result can be saved and assembled into a story. A story can then be shared via a public link. The inclusion of compose functionality makes DIVE an authoring and publishing tool. DIVE allows users to easily analyse data and share their findings, though no security (i.e., authorisation and authentication) is enforced for shared stories.

3. Extending the Concepts Introduced in DIVE in Learning Analytics

The three main innovations introduced in DIVE include data model inference, mixed-initiative visualisation recommendation, and the generation of sharable analytic narratives. All three are very relevant to learning analytics. The automated data model inference and mixed-initiative visualisation recommendation in particular would allow prototype dashboards and applications to be built in co-design sessions. Educators participating in co-design sessions would be able to select or suggest the variables of interest and the design facilitator or even the educator, or a support member of the teaching team could see a preview in real time. This nicely addresses the challenges of time and technical skill in progressing through an application design process. There are however opportunities to further enhance DIVE to better suit the learning analytics domain.

3.1 Data Ingestion and Inspection

DIVE uses a heuristic algorithm to determine the semantic categorization of columns (i.e., fields) in a tabular dataset. The current algorithm would need to be tested with clickstream data (originating from learning management systems and embedded tools such as Echo360), student demographic, enrolment and assessment data. DIVE associates a single tabular dataset with a project. The data model would need to be extended for learning analytics to allow multiple related datasets. This would allow for datasets from different systems to be uploaded and linked via a student identifier.
Within one-on-one sessions with an educator or multiple participant co-design sessions, real student datasets may not be available or unable to be used due to privacy concerns. In order for the tool to be useful within either setting, functionality is required to be able to synthesise data for various related student datasets. It is envisioned that the fields for dummy datasets would need to be entered along with the parameter configurations (e.g., max, min and standard deviation). An additional requirement would be to allow for student sub-groups to be modelled as this would allow for discussion around the interpretation of visualisations and provision of personalised feedback within co-design sessions. Data synthesis is both under researched and utilised within learning analytics research but ideas from recent open source data synthesize tools such as Synner (Mannino & Abouzied, 2020) can be incorporated within DIVE. Synner provides a declarative user-interface for users to describe the statistical distribution of fields and the relationship between fields.

3.2 Visualisation and Drilldown

In recent years many visualisation recommendation algorithms have been proposed (Wongsuphasawat et. al., 2020, Lima et. al., 2020) and the algorithm within DIVE can be updated to incorporate ideas from recent research. A key feature of the visualisation engine embedded within DIVE is that mixed-initiative recommendations are supported. The mixed-initiative feature is important because unlike other recommendation engines that automatically select the variables to be displayed in a visualisation, DIVE allows the user to specify the fields they are interested in. Within a co-design session, this feature will facilitate the inclusion of feedback from the educator thereby directing the visualisations to be generated.

3.3 Story Creation and Sharing

The Compose feature in DIVE allows visualisations to be collated and annotated on a web page. While this is a simple idea, it essentially completes the full workflow from design to a deployed analytics artefact. Imagine the power of generating a custom course-specific dashboard for an educator in a co-design session and creating a working analytics tool that can be re-used with real data for the forthcoming semester. However, an essential prerequisite for implementing DIVE within a university setting is password protection of shared analytics which would need to involve integration with university single sign on to address privacy issues relating to student demographic and assessment data.

4. Addressing the “Clicks to Construct” Issue in Learning Analytics

The automated recommendation of time series charts however will not necessarily provide the insight that educators need to adapt courses and provide personalised feedback. While trace data (also known as clickstream data) has proven difficult to map to higher level constructs (i.e., the “clicks to construct” issue) (Knight and Shum, 2017), recent learning analytics research has focussed on addressing the issue by adding linked data or metadata to the events captured within a clickstream.

In this paper, we use the term “semantic layer” to describe the inclusion of linked or metadata to develop meaningful constructs for visualisations and metrics. DIVE’s data model would need to be extended to allow for additional linked and metadata to be uploaded or synthesised (i.e., dummy data). It is envisaged that initially three semantic layers that use techniques proposed from recent learning analytics research (i.e., 2019) could be added to the DIVE data model. Two (2) of the techniques require additional metadata (e.g., educator annotations of course material to scheduled delivery week and estimates of activity duration) while the additional one (1) proposes an alternate way of processing student assessment data to generate a sequential Sankey diagram.

4.1 Classifying Student Activity as Ahead, Preparing, Catching Up and Revising

In Gašević et, al., (2019), a novel algorithm to classify student interaction within blended or fully
online course material is proposed. Clickstream data is enriched by linking events to content modules and their scheduled delivery date. The additional metadata allows student access to be classified as being ahead, preparing, catching up, revisiting or not accessing the course (introduced to provide a 100% student population representation). The classification can be visualised (see Figure 3) to provide a holistic weekly overview of a student cohort’s progress with coursework. A Sankey visualisation can also be used to display student transitions through semester weeks. 

4.2 Comparing Activity Time Estimates to Actual Student Completion Time

Ginda et al., (2019) enriches data relating to course access events with the course section being completed, the average time it has taken students to complete, and an estimate of activity completion time provided by the course designer. The resulting visualization (see Figure 4), allows educators to easily identify sections and activities that are either completed too quickly or slowly thereby increasing the estimated time for students to complete the module.

4.3 Analyse Student Grade Transitions

The grade Sankey visualization introduced by Deng et al., (2019) provides new insight on progression and performance in course assessment. The grade Sankey is created by calculating how students transition between grades for each assessment item and then displaying the transition flow on a sequential Sankey diagram with fixed axis labels for each assessment item. The visualization provides a way to identify and provide feedback and support to students with decreasing grades or notable changes in performance.

Figure 3: Holistic overview of progress made by cohort of students in completing online content.
Figure 4: Difference between activity time estimates and actual time spent by students.

Figure 5: Sequential Sankey used to illustrate student grade transitions.
5. Use Cases within Learning Analytics

In this section we present two (2) example use cases for using DIVE after enhancements for learning analytics have been implemented within the tool. The examples serve to showcase the potential of the key innovations introduced in DIVE and demonstrate that the proposed workflow is flexible enough to be used in a variety of learning analytics design sessions.

5.1 Educational Data Scientist Works with Teaching Team in a Co-Design Session

A teaching team consisting of a course coordinator, lecturers teaching sections of the course and the faculty assigned learning designer has just completed restructuring a course for flipped delivery (i.e. online modules that need to be completed prior to active face to face workshops). The course coordinator, nervous about delivering the course for the first time, contacts the central learning analytics unit at the university. The university while investing in commercial analytics products has also placed an emphasis on developing custom co-designed solutions to better meet the analytics needs of specific domains. The central learning analytics unit proposes a co-design session with the faculty assigned educational data scientist.

The co-design session begins with the educational data scientist asking questions about the course redesign. No real student data has been collected as the course has not yet run, but the educational data scientist is able to synthesize mock (i.e., dummy) data for each assessment item and the student event clickstream for each of the online modules within DIVE. The educational data scientist then clicks on the visualisations tab and asks the team what questions they would like to answer. The lecturers would like to know whether students are completing the online module before they come to campus so that they can adapt the weekly face to face workshops accordingly. The educational data scientist then uses the mixed-initiative visualisation recommendation engine by selecting the variables of interest in the clickstream and applies a semantic layer that includes information on whether students have completed the online module in the scheduled delivery week. The recommended visualisations include a time series plot of access, a bar chart of the % of students that accessed content per teaching week, a box plot chart showing the percentage of students that have completed each module, and a stacked bar chart showing the percentage of students that were categorised each week according to whether they were on track, ahead, catching up or revising content. The teaching team all agree that the stacked bar charts and the box plot both provide a good overview of the whole cohort. The educational data scientist saves the stacked bar chart and the box plot and also informs the team that in order to get these meaningful charts they will need to map each module from the LMS to a teaching week. The educational data scientist repeats the process of asking questions, specifying variables of interest and saving recommended visualisations as the co-design session progresses.

At the conclusion of the co-design session, all of the visualisations are added to a dashboard that is then shared with the team. The educational designer informs the team that the dashboard will automatically be populated with real student data once the semester begins.

5.2 Educator Created Narrative Visualisation

An educator would like to make a case for redesigning the assessments and their weightings in a course. The educator has a sense that students who get a good grade in early assessment items are performing badly in the final exam. The educator has a download of the Grade Centre file from the Blackboard LMS. Viewing the data in Excel does not directly reveal any insights. The educator is aware of a new tool called DIVE being piloted at the university. The educator logs into the tool and uploads the gradebook file. DIVE is able to automatically detect assessment scores and matches these to a semantic layer that can calculate the grade transitions of each student. On the visualisation recommendation page, the various histograms are proposed including histograms of scores for each assessment, boxplots for each assessment on a single chart, and a sequential Sankey diagram. The educator thinks that the sequential Sankey is able to illustrate clearly that the final assessment task is
either too difficult or incorrectly weighted and proceeds to add the visualisation to a story. The educator annotates the visualisation and saves a version for inclusion in a presentation for the course review panel.

6. Discussion

After reviewing DIVE, suggesting enhancements specifically for learning analytics and addressing the ‘clicks to construct’ issue in learning analytics, we are now able to address the key research questions.

1. Is it possible to produce working prototypes within a design session between a learning analytics practitioner and educators?
   From our review of the functionality of DIVE it would seem that the platform could facilitate the development of working prototypes for educators, however enhancements would be needed to make the platform more suitable to the context of learning analytics. The workflow proposed by DIVE (i.e., Ingest and Inspect, Visualise, Analyse and Compose) simplifies data exploration and data story authoring without requiring specialist skill sets (i.e. data science or programming skills). Key enhancements are however required to contextualise DIVE for use in learning analytic co-design sessions including data model inference for clickstream and assessment data and the ability to synthesise dummy data. Mixed-initiative visualisation recommendation was also preferable instead of a simple gallery of visualisations, where the variables of interest could be specified, which would allow co-design session participants to direct the recommended visualisations.

2. Can learning analytic applications be automatically generated?
   The key innovations that facilitated the automatic generation of visualisations within DIVE were the data inference model and the visualisation recommendation engine. The data inference model was able to classify fields within the uploaded dataset which was then used by the visualisation recommendation engine. The visualisation recommendation engine used the field level classifications to generate appropriate variable encodings for representation on charts and visualisations. Incorporating the ability to generate charts directly from uploaded data is essentially what provides the functionality to automatically generate an application without the underlying programming knowledge. Visualisation recommendation is an active research area and requires further investigation by the learning analytics research community.

3. Can the generated learning analytics application address the ‘clicks to construct’ issue?
   The current functionality and workflow provided by DIVE would be directly applicable to most domain areas, however some enhancements would be required to address this issue more specifically. One of the challenges with learning analytics is that trace data of discrete student events with a learning system provides very little insights into the actual learning of students. In order to address this issue latest research on higher level constructs needs to be incorporated into the data model inference and aggregate analysis modules in DIVE. Higher level constructs are usually created by including a semantic layer mapping discrete clickstream events to additional metadata. Within this paper we proposed an initial set of three (3) layers, with the aim of increasing this list as new research emerges.

7. Conclusion and Future Directions

The DIVE application has served as a good example of how a modern data exploration workflow tool can be reviewed and adapted to support learning analytics workflows. The innovations introduced within DIVE include data model inference, mixed-initiative visualisation recommendation and the ability to assemble visualisation into shareable narrative stories. These innovations coupled with the adaptations proposed in this paper, have the ability to transform the current design, development and deployment processes currently used in learning analytics.

Future research involves implementing the enhancements within DIVE and conducting a series of case studies with practitioners, educators, learning designers, educational data scientists and data engineers to evaluate the processes discussed here in practice. It is hoped that there will also be
interest from the learning analytics research community to further explore research in visualisation recommendation and data story creation.

References


A Unified Approach for Analysing Computer Science Courses: An Australian Case Study

Sameera JAYARATNA*a, Caslon CHUAa, Mohommed Eunus ALIb & Timos SELLIS*
aSwinburne University of Technology, Australia
bBangladesh University of Engineering and Technology, Bangladesh
*mjayaratna@swin.edu.au

Abstract: A large variety of Computer Science (CS) and Information and Communications Technology (ICT) programs are offered in different institutions across Australia. Even the same institution has several majors of CS programs due to its recent popularity among students and demand in the job market. Current practices in CS education lack a unified approach to analyse and compare these programs, and consequently cannot assess the students in different programs based on a common platform. In this paper, we propose a unified and systematic approach to analyse CS units and courses. Our approach is based on CS curriculum guidelines, according to which each subject (or unit) of a program (or course) can be mapped to knowledge areas. The insights gained through the proposed unified approach could help education providers, students, and governing bodies to understand the fundamentals of CS and ICT programs. A case study involving data from CS programs offered by two Australian higher education institutions show the efficacy of our proposed approach.

Keywords: Computer science education, course analysis, course comparison

1. Introduction

Computer Science education is increasingly becoming popular across the globe. According to the Australian Computer Society (ACS), the annual enrolments in Computer Science (CS)/Information and Communications Technology (ICT) related programs at universities in Australia are 50 percent higher than a decade ago (ACS & Deloitte Access Economics, 2019). With the increasing demand, most higher education providers offer variants of CS/ICT programs. Currently, around 40 education providers, with a total of 340 CS/ICT programs, hold ACS accreditation (Grayston, 2019).

A course refers to a program of study such as Bachelor of CS or Bachelor of ICT. Courses offered in both undergraduate and postgraduate levels by different institutions tend to differ in terms of structure, the number of units taught, and the topics covered. A unit refers to any subject such as Introduction to Programming or Software Engineering, offered under a course. It is unlikely to encounter courses in two different institutions that are identical in terms of the units they are composed of. Even the units with similar names offered by different institutions may not be similar in terms of the topic coverage (the number of topics and the breadth of each topic covered in the unit). Thus, a significant variation can be observed among courses in the same discipline. Also, CS curricula change over time due to the rapid change of technology.

The ability to effectively analyse multiple courses is paramount to all the participants of CS education: education providers, prospective students, current students, industry professionals, and professional bodies. For education providers, course analysis is vital to assess their courses according to design guidelines and compare against other similar courses. The prospective students can benefit from an effective comparison of the differences in CS programs when making choices. Students who are currently enrolled in a CS program can benefit from an effective analysis of units offered in the program. For instance, an effective comparison of units available as elective units of the program can be vital for a student’s overall performance. The industry professionals’ ability to understand how a CS program would cater to their requirements may enable the selection of the most suitable graduates. Professional bodies such as ACS can gain insights from effective course analysis to ensure the quality of CS courses offered by education providers. Current practices in CS education lack a unified approach
to analyse and compare CS programs. Consequently, cannot assess the students in different programs uniformly. To address the above issues, we propose a unified approach to analyse CS courses and units.

Our approach is motivated by the curriculum guidelines for CS programs published by the Association for Computing Machinery (ACM), IEEE Computer Society, and ACS. These curriculum guidelines fundamentally use two components to describe the content of a CS unit: Body of Knowledge (BOK) and Learning Taxonomy. ACM and IEEE jointly published curriculum guidelines (Sahami et al., 2013), which discuss the BOK relevant for undergraduate programs in computing. Similarly, ACS published Core Body of Knowledge (CBOK) for ICT professionals (ACS, 2015). These curriculum guidelines recommend a unit to cover the knowledge areas defined in a BOK at appropriate levels of difficulty (or mastery) defined by a learning taxonomy. These organisations have also recommended learning taxonomies, like Bloom’s taxonomy (Bloom, 1956) and Bigg’s SOLO taxonomy (Biggs & Collis, 2014).

Most CS programs are designed according to a set of curriculum guidelines, such as those stated above. Therefore, the coverage of a unit can be mapped to a BOK and a learning taxonomy. In this study, we present a unified representation of units that can be adopted for any given BOK and learning taxonomy combination. For instance, Australian CS programs follow ACS guidelines, which involve ACS CBOK and Bloom’s taxonomy. Thus, the proposed approach enables all CS units and courses that follow ACS guidelines to be uniformly assessed using ACS CBOK and Bloom’s taxonomy.

In this paper, we propose a unified approach for analysing and comparing CS units and courses mapped to a BOK with respective levels of difficulty defined in a learning taxonomy. First, we explain two fundamental elements, namely: BOK and learning taxonomy, in representing the units. Then, we explore how the stakeholders of CS education, such as educators and students, can gain valuable insights on units and courses using the proposed unit representation. We demonstrate the applicability of the proposed approach via an Australian case study. CS programs in Australia follow ACS curriculum guidelines based on ACS CBOK and Bloom’s taxonomy. Therefore, we adopt the proposed approach to the Australian context by using the ACS CBOK and Bloom’s taxonomy as the instances of BOK and learning taxonomy, respectively. The dataset used is generated by universities as part of the ACS course accreditation process. The availability and the quality of the dataset facilitate the immediate adoption of the proposed approach in any institution that follows ACS guidelines. The contribution of this paper can be highlighted in three aspects: (i) we represent CS units based on knowledge area coverage, (ii) we propose a novel unified approach to analyse CS courses and units based on the unified unit representation, and (iii) we present a case study with a dataset obtained from Australian universities to show the applicability of the proposed approach.

The next section briefly reviews the related work. Then, Section 3 explains how units are represented based on BOK and learning taxonomy. Section 4 presents the case study, which demonstrates how the unified representation discussed in Section 3 can be used to analyse courses and units. Finally, Section 5 draws conclusions from the case study performed and discusses the potential extensions of the proposed approach in future work.

2. Related Work

Many studies have discussed methodologies for analysing CS course curricula to provide insights to educational providers (Méndez, Ochoa, & Chiluiza, 2014; Oliver, Dobele, Greber, & Roberts, 2004; Pedroni, Oriol, & Meyer, 2007; Sekiya, Matsuda, & Yamaguchi, 2015). Méndez et al. (2014) presented a set of techniques that relies on historical academic data for various applications like estimating course difficulty and identifying dropout paths. The results were then used to gather recommendations for curriculum design. Pedroni et al. (2007) proposed a more systematic approach to define courses intuitively. However, their approach requires teachers to define courses with units of knowledge, which takes a lot of time and effort. Sekiya et al. (2015) compare CS-related courses offered across universities using natural language processing to analyse how a course curriculum is distributed over knowledge areas defined by ACM & IEEE. However, it does not capture the depth in which a course covers the knowledge areas. In contrast, Oliver et al. (2004) analyse the cognitive difficulty of different courses offered by a university using Bloom’s taxonomy while ignoring the aspect of knowledge areas.

Learning taxonomies such as Bloom’s taxonomy and Bigg’s SOLO taxonomy are used in computing education (Fuller et al., 2007; Shuhidan, Hamilton, & D'Souza, 2009). They are often used
to indicate different stages of learning development, which is useful to distinguish the appropriateness of learning outcomes at different levels in courses. Bloom’s taxonomy appears to be the most widely used taxonomy to state learning goals in computing studies (Johnson & Fuller, 2006; Masapanta-Carrión & Velázquez-Iturbide, 2018). This taxonomy is used as a reference for the classification of knowledge and competencies to be acquired via an education module. Bloom’s taxonomy has also inspired studies on course comparison in areas of education other than CS (Hoffmann, 2008).

3. Unified Representation of Units

In this section, we first describe the fundamental elements used to describe the content of a CS unit: BOK and Learning Taxonomy. Then, we discuss how to represent a CS unit based on these elements.

3.1 Body of Knowledge and Learning Taxonomy

Multiple organisations have produced documents describing BOK for CS-related disciplines. IEEE computer society published Software Engineering Body of Knowledge (SWEBOK) (Bourque & Fairley, 2014), while ACM and IEEE Computer Society jointly published curriculum guidelines (Sekiya et al., 2015) discussing the BOK relevant for undergraduate programs in computing. Similarly, ACS published CBOK for ICT professionals (ACS, 2015).

The ACM and IEEE computer society’s document on curriculum guidelines (Sahami et al., 2013) has organized the BOK into a set of 18 knowledge areas corresponding to topical areas in computing. These knowledge areas include software engineering, operating systems, and programming languages (Sekiya et al., 2015). Each knowledge area is further organized into a set of knowledge units. The guidelines illustrate how knowledge areas may be covered and combined in courses. As guidance, each knowledge unit lists a set of topics and learning outcomes students are expected to achieve. Each learning outcome is associated with a level of difficulty. This document defines three difficulty levels based mainly on Bloom’s taxonomy. The levels of difficulty are defined as familiarity, usage, and assessment. In summary, these curriculum guidelines provide a framework for universities to map the courses they offer to either knowledge areas or knowledge units with the respective levels of difficulty.

ACS provides guidelines for course design using the CBOK for ICT professionals (ACS, 2015). ACS has incorporated these guidelines into their course accreditation process to ensure the quality of computing education in Australia (ACS, 2020). The CBOK introduces 5 knowledge areas: ICT Problem Solving, ICT Professional Knowledge, Technology Resources, Technology Building, and ICT Management (ACS, 2015), which are further divided into topics. ACS accreditation process requires the coverage of these knowledge areas by each unit of a course to be captured according to the ACS guidelines. Like ACM, ACS mainly utilises Bloom’s taxonomy to indicate levels of difficulty (ACS, 2020). ACS also allows the use of alternative indicators such as Biggs SOLO taxonomy. As per the ACS guidelines, a unit is designed such that each unit activity covers a set of knowledge areas at specific levels of difficulty. Unit activities may include lectures, tutorials, and exams. In summary, ACS provides a framework to map a unit’s activities to knowledge areas with respective levels of difficulty.

The main difference between ACS guidelines and ACM guidelines is the structure of the BOKs. In addition, ACS guidelines capture a unit’s coverage in terms of unit activities while ACM guidelines capture the same in terms of a unit’s learning outcomes. Unit activities and learning outcomes may be considered as two types of unit components. Thus, in summary, both guidelines map unit components (learning outcomes or activities) to a BOK with respective levels of difficulty defined in a learning taxonomy. Accordingly, we describe a unified representation of units in the next section.

3.2 Unit Representation

Units can be considered as the building blocks of a course. Therefore, we first investigate how the coverage of a unit is presented against a BOK, which defines a set of Knowledge Areas (KA). A unit can be described using a set of unit components, which may be either a set of learning outcomes or a set of activities (e.g., tutorials, assignments, exams). Each component of a unit can be mapped to the
defined KAs with respective levels of difficulty defined in a learning taxonomy. Each component of a unit covers one or more KAs.

In this study, a unit is represented based on the summarised levels of difficulty across all the unit components per KA. A unit’s overall coverage of a KA is indicated by the Difficulty Rating of that KA. Given a set of $n$ KAs and a unit $X$ with $c$ components, the Difficulty Rating of the $i^{th}$ KA, $DR_{i,X}$ is calculated using the level of difficulty at which each component of unit $X$ covers that KA.

$$DR_{i,X} = \text{median}(KA_{i,X,1}, KA_{i,X,2}, \ldots, KA_{i,X,c})$$

where $KA_{i,X,c}$ refers to the difficulty level at which the $c^{th}$ component of unit $X$ covers the $i^{th}$ KA. Only the components that evaluate a given KA are used to calculate the median. In the case of an even number of components, the lower of the two middle values will be taken as the median, because difficulty levels defined in learning taxonomies are of ordinal scale. Then, unit $X$ is presented as an $n$-tuple.

$$\text{Unit}_X = < DR_{1,X}, DR_{2,X}, \ldots, DR_{n,X} >$$

Table 1 presents the coverage of a sample unit $Q$ with three components against a given BOK with five KAs. Assume the learning taxonomy used defines levels of difficulty as $L1$, $L2$, and $L3$, where $L1$ is the least difficult and $L3$ the most. The value in each cell indicates the level of difficulty at which the respective component evaluates a KA. An empty cell indicates that the component does not cover the respective KA. Based on Eq. 1, Table 1 presents the Difficulty Ratings of unit $Q$ in the last column. Then, unit $Q$ can be represented as a 5-tuple: $< L3, L3, L2, L1, L2 >$ according to Eq. 2.

<table>
<thead>
<tr>
<th>Knowledge Area</th>
<th>Unit Components</th>
<th>Difficulty Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>$KA_1$</td>
<td>$L1$ $L3$</td>
<td>$L3$</td>
</tr>
<tr>
<td>$KA_2$</td>
<td>$L2$ $L3$ $L3$</td>
<td>$L3$</td>
</tr>
<tr>
<td>$KA_3$</td>
<td>$L2$ $L2$</td>
<td>$L2$</td>
</tr>
<tr>
<td>$KA_4$</td>
<td>$L1$ $L1$</td>
<td>$L1$</td>
</tr>
<tr>
<td>$KA_5$</td>
<td>$L3$ $L3$ $L2$</td>
<td>$L3$</td>
</tr>
</tbody>
</table>

4. Case Study

In this paper, we present a case study set in the Australian context. Therefore, we implement the unit representation defined above based on the data from Australian universities. As mentioned in Section 3.1, Australian universities follow the ACS course accreditation guidelines to design their computing courses. The accreditation process requires a university to submit a document indicating how units of a CS course covers the CBOK. Therefore, each CS program that applies for ACS accreditation maps each unit’s components to KAs defined in the CBOK with respective levels of difficulty. ACS uses Bloom’s levels for indicating the levels of difficulty. In this section, we first describe two fundamental elements: ACS CBOK and Bloom’s levels, which indicate levels of difficulty. Then, based on the CBOK and Bloom’s level, we represent units used for measuring and comparing courses.

4.1 ACS Core Body of Knowledge

In this paper, we discuss the specifics of ACS CBOK, because while our work can be applied to CS courses based on any BOK, it is primarily focused on the Australian context. ACS accreditation guidelines indicate 5 knowledge areas which are further divided into 19 topics. In this paper, we refer to the 5 knowledge areas as ‘Categories’ and 19 topics as ‘Knowledge Areas (KA)’ to improve readability. Table 2 presents the categories and KAs under each category. An Australian university program should provide appropriate coverage of the KAs defined by the ACS CBOK to receive course accreditation (ACS, 2015). A program consists of multiple units, and a unit is designed with multiple
activities (e.g., assignments, tutorials, exams). Each activity assesses one or more KA at varying levels of difficulty. ACS recommends Bloom’s levels as the indicator of difficulty levels.

Table 2. ACS CBOK Knowledge Areas

<table>
<thead>
<tr>
<th>Category</th>
<th>Knowledge Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICT Problem Solving</td>
<td>1. Abstraction</td>
</tr>
<tr>
<td></td>
<td>2. Design</td>
</tr>
<tr>
<td>ICT Professional Knowledge</td>
<td>3. Ethics</td>
</tr>
<tr>
<td></td>
<td>4. Professional expectations</td>
</tr>
<tr>
<td></td>
<td>5. Teamwork concepts &amp; issues</td>
</tr>
<tr>
<td></td>
<td>6. Interpersonal communications</td>
</tr>
<tr>
<td></td>
<td>7. Societal issues</td>
</tr>
<tr>
<td></td>
<td>8. Understand of ICT profession</td>
</tr>
<tr>
<td>Technology Resources</td>
<td>9. Hardware &amp; software fundamentals</td>
</tr>
<tr>
<td></td>
<td>10. Data &amp; information management</td>
</tr>
<tr>
<td></td>
<td>11. Networking</td>
</tr>
<tr>
<td>Technology Building</td>
<td>12. Programming</td>
</tr>
<tr>
<td></td>
<td>13. Human factors</td>
</tr>
<tr>
<td></td>
<td>14. Systems development</td>
</tr>
<tr>
<td></td>
<td>15. Systems acquisition</td>
</tr>
<tr>
<td>ICT Management</td>
<td>16. IT governance &amp; organisational issues</td>
</tr>
<tr>
<td></td>
<td>17. IT project management</td>
</tr>
<tr>
<td></td>
<td>18. Service management</td>
</tr>
<tr>
<td></td>
<td>19. Security management</td>
</tr>
</tbody>
</table>

4.2 Bloom’s Taxonomy

Bloom’s taxonomy is a well-known learning taxonomy from pedagogy. The learning taxonomy devised by Bloom divides the cognitive aspects of learning into six hierarchical levels, starting from the simplest to the most complex. Here, ‘the levels’ indicate the levels of difficulty. That is, the first ones must typically be mastered before the next one can take place. Bloom’s levels are presented in Table 3. The model was revised in 2001 by Anderson and a team of cognitive psychologists (Anderson et al., 2001). Bloom’s taxonomy appears to be the most widely used taxonomy to state learning goals in computing studies (Johnson & Fuller, 2006; Masapanta-Carrión & Velázquez-Iturbide, 2018). Even though there are concerns about the appropriateness of Bloom’s taxonomy in CS education, they are mostly about the difficulties of usage (Masapanta-Carrión & Velázquez-Iturbide, 2018). Studies also focus on ways to improve the understanding of educators to ensure the effective use of Bloom’s taxonomy (Richard, Judy, Raymond, Simon, & Sabina, 2013; Starr, Manaris, & Stalvey, 2008).

Table 3. Bloom’s Taxonomy

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Knowledge</td>
</tr>
<tr>
<td>2</td>
<td>Comprehension</td>
</tr>
<tr>
<td>3</td>
<td>Application</td>
</tr>
<tr>
<td>4</td>
<td>Analysis</td>
</tr>
<tr>
<td>5</td>
<td>Synthesis</td>
</tr>
<tr>
<td>6</td>
<td>Evaluation</td>
</tr>
<tr>
<td></td>
<td>Recall facts and basic concepts</td>
</tr>
<tr>
<td></td>
<td>Explain ideas or concepts</td>
</tr>
<tr>
<td></td>
<td>Use information in new situations</td>
</tr>
<tr>
<td></td>
<td>Draw connections among ideas</td>
</tr>
<tr>
<td></td>
<td>Justify a stand or decision</td>
</tr>
<tr>
<td></td>
<td>Produce new or original work</td>
</tr>
</tbody>
</table>

4.3 Data Set

The data analysed in this article are collected from two higher education institutions in Australia. The composition of the data set is shown in Table 4. Institution A offers a Bachelor of Computer Science course with multiple majors, including Cybersecurity, Datascience, Networking, and Software
development. Curriculum for each major (A1-A4) consists of 8 core units, 8 major-specific units, and 8 electives. The core units are common to all the majors. Bachelor of Networking course from institution B (B1) consists of 21 core units and 3 electives. Like Institution A, Institution B allows the electives to be either CS or non-CS units. Non-CS units (e.g., Introduction to Management) offered by non-CS departments do not have ACS CBOK mapping. However, the enrollments of such units can be considered low compared to the enrollments of CS units by the students. The units in the dataset use multiple activities (components) to assess the knowledge of CS students. Table 5 presents one of the units used in this study. The value in each cell indicates the Bloom’s level at which the respective activity evaluates a KA. An empty cell indicates that the activity does not cover the respective KA.

Table 4. Data Set

<table>
<thead>
<tr>
<th>ID</th>
<th>Institution</th>
<th>Course Name</th>
<th>Major</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>Institution A</td>
<td>Bachelor of Computer Science (UG-CS1)</td>
<td>Cybersecurity</td>
</tr>
<tr>
<td>A2</td>
<td></td>
<td></td>
<td>Data science</td>
</tr>
<tr>
<td>A3</td>
<td></td>
<td></td>
<td>Networking</td>
</tr>
<tr>
<td>A4</td>
<td></td>
<td></td>
<td>Software development</td>
</tr>
<tr>
<td>B1</td>
<td>Institution B</td>
<td>Bachelor of Networking (UG-CS2)</td>
<td></td>
</tr>
</tbody>
</table>

Table 5. Coverage and Difficulty Rating of a Unit

<table>
<thead>
<tr>
<th>Knowledge Areas</th>
<th>Lectures</th>
<th>Assignment</th>
<th>Tutorials</th>
<th>Exam</th>
<th>Difficulty Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstraction</td>
<td></td>
<td></td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Design</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Ethics</td>
<td></td>
<td></td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Prof. expectations</td>
<td></td>
<td></td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Teamwork concepts</td>
<td>2</td>
<td></td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Interpersonal comm.</td>
<td></td>
<td></td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Societal issues</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Understand of ICT prof.</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Hardware &amp; software</td>
<td></td>
<td></td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Data &amp; info. management</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Networking</td>
<td></td>
<td></td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Programming</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Human factors</td>
<td></td>
<td></td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Systems development</td>
<td></td>
<td></td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Systems acquisition</td>
<td></td>
<td></td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>IT gov. &amp; org. issues</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>IT project management</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Service management</td>
<td></td>
<td></td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Security management</td>
<td></td>
<td></td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

4.4 Unit Analysis

The units offered by a CS program tend to cover combinations of KAs at various levels of difficulty. A unit would typically have one or more common KAs with another unit of the same CS program. However, a unit description is often textual and includes topics covered, learning outcomes, and assessment tasks. The comparison of textual descriptions to understand similarities or dissimilarities between units can be tedious and time-consuming. In this section, we demonstrate unit analysis from two perspectives: a current student’s perspective and an education provider’s perspective.

First, we look at the scenario of a current student. A student is currently studying a CS program with several elective units. She intends to select either Programming Fundamentals unit or Mobile Programming unit as one of the electives. She needs to understand the differences between the units to make an informed selection. Figure 1 presents the comparison of units based on their coverage of KAs. As per Figure 1, Programming Fundamentals unit covers a larger number of KAs (11) than Mobile Programming (7). More specifically, the Mobile Programming unit covers the Networking KA while the other unit does not. Also, activities of Mobile Programming unit assess students in Data & information management KA at a higher level (4) than that of Programming Fundamentals unit (2). Accordingly, the student may select a unit that aligns with her preferences.
Next, we consider a scenario of an education provider. An education provider is offering four majors under a CS program. Course designers are expected to update the program for which they need to review the current course composition. Figure 2 shows the spread of all units offered as core units, which are common to all majors and major-specific units for each major (A1-A4) over 19 KAs. The values of cells indicate the Difficulty Rating of units calculated as per Section 3.2. The values are indicated in a colour scale ranging from light blue to dark blue as they vary from 1 to 6. This visualization helps the course designers gain insights into KA coverage of units. For example, one can...
observe that almost all the units cover the Problem Solving KA category. In addition, Security Management KA is covered not only by security-related units but also by networking related units. Interestingly, there is a lack of units that cover Service Management KA. Such insights may help course designers identify gaps in the program that can be filled by introducing new units.

4.5 Course Analysis

Systematic representation of courses can be beneficial for many parties, including education providers, prospective students, and professional bodies. For instance, universities can gain insights into courses offered by them. Prospective students get the ability to make an informed decision when choosing a course. The overwhelming process of browsing numerous course curricula may be avoided by using a summary of vital information. Additionally, professional bodies such as ACS can gain insights that can contribute to shaping the overall landscape of CS education by assessing all the courses through a systematic representation. In this section, we demonstrate the benefits of course analysis from a prospective student’s perspective and an employer’s perspective. For each perspective, we illustrate a scenario based on the dataset shown in Table 4.

First, we look at the scenario of a prospective student. A student interested in studying computing has selected a university which offers multiple majors under its Bachelor of Computer Science program. The student has a preliminary understanding of the domains that would provide future job opportunities. Accordingly, he is interested in Data science, Cybersecurity, and Software development majors. He needs to understand the differences between the majors to select the suitable major for him. Figure 3 uses a box and whisker plot to summarise the units of each major across the 19 KAs. Each major is analysed based on the set of core and major-specific units. This plot indicates the minimum, maximum, and median Difficulty Ratings of each major. Figure 3 indicates that the majors are relatively different in terms of their coverage of KAs. This student is particularly interested in the Security management KA and Programming KA of the majors. He would like to gain significant exposure to Security management KA and extensive exposure to Programming KA. As per Figure 3, Security management KA is covered by each of the three majors. Units of Data science and Cybersecurity majors are concentrated on Difficulty Rating 3 in Security management KA. However, Figure 2 indicates that in contrast to the Cybersecurity major, the Data science major does not cover Security management KA in any major-specific units. The units of Software development major are distributed between Difficulty Ratings of 1 and 4 in Security management KA. Also, Figure 2 indicates that the Software development major includes a few major-specific units that cover Security management KA. As per the analysis, the Software development major offers significant exposure to Security management KA while the Cybersecurity major provides extensive exposure. Furthermore, Figure 3 indicates that Programming KA is covered by each major, and 50% of units in each major have a Difficulty Rating of 3 or 4. Also, units of Data science major and Cybersecurity major appear to be similarly distributed in Programming KA with a minimum Difficulty Rating of 3. In contrast, the minimum Difficulty Rating within Software Development major units is 2. According to Figure 2, most of the units in Software Development major cover Programming KA. Therefore, it is evident that the Software Development major would offer the most exposure to Programming KA in comparison to the other two majors. As per the above findings, the Software development major would best cater to the student’s requirements regarding Security management KA and Programming KA.

Next, we look at the scenario of an employer. An industry professional is looking to employ networking graduates according to specific criteria. She wants to find the CS program that is most likely to produce graduates fitting her criteria. Programs under consideration are UG-CS2 program of Institution B and UG-CS1 program of Institution A. The KAs of most interest in her criteria include Networking and Interpersonal communication. The industry professional can utilise Figure 4 to make an informed choice between the programs. Significant differences can be observed between the courses, even though both are specialised in networking. Regarding the coverage of Networking KA, the maximum Difficulty Rating of UG-CS2 is higher at 6, while the value of Institute A’s UG-CS1 is only at 4. Also, the minimum Difficulty Rating of UG-CS2 is higher at 3, while the value of UG-CS1 is only at 2. Also, the inter-quartile range of UG-CS2 is larger than that of UG-CS1. Therefore, we can observe that the coverage of Networking KA is consistently higher in Institute B’s UG-CS2 than in Institute A’s UG-CS1. The minimum (2) and maximum (5) Difficulty Ratings of Interpersonal communication KA are similar in both programs. However, 50% of the units in UG-CS2 cover the Interpersonal
communication KA at Difficulty Ratings of 3 or 4. Accordingly, the employer may consider the UG-CS2 program better with regards to Networking KA and Interpersonal communication KA. Similarly, the employer can compare the two programs based on their coverage of KAs relevant to her criteria.

In summary, the course analysis showed the ability to capture the contrast in CS programs through visualizations. It demonstrated how stakeholders, such as students and industry professionals, can explore a set of CS programs for a variety of purposes.

![Figure 3. Comparison of majors.](image1)

![Figure 4. Comparison of networking courses across institutions A and B.](image2)

5. Conclusion and Future Work

We proposed a unified approach for analysing CS courses based on a BOK and a learning taxonomy. The case study based on the Australian context demonstrated the adaptability of the proposed approach for any specific context. Thus, ACS CBOK and Bloom’s taxonomy was used as the BOK and the learning taxonomy of the unit representation, respectively. The use of ACS CBOK and Bloom’s taxonomy in the ACS accreditation process facilitates the immediate adoption of the proposed approach. The case study demonstrated the use of the proposed approach for unit and course analysis. The proposed unified unit representation has multifaceted benefits. For example, education providers can analyse the courses objectively, which will help them improve/redesign existing courses; professional bodies such as ACS can compare different programs and assess their strengths and weaknesses. Furthermore, a comparison of courses across different universities and courses with different majors

233
can provide multiple perspectives that can be beneficial to educators, students, and industry professionals. In addition, unified unit representation enables effective unit comparison, which overcomes the limitations inherent to textual description comparison.

Though we present a preliminary study based on our unified approach, it opens several avenues for future works. In the future, we intend to extend this approach to assess student performance in each KA. Students may be represented in a unified manner based on all the units completed under a course. This representation may indicate the variation of KA coverage among students due to the choice of elective units. Also, we plan to analyse a larger number of courses to get a broader perspective on CS education in Australia.

References

Identifying Student Engagement and Performance from Reading Behaviors in Open eBook Assessment

Brendan FLANAGAN*\(^a\), Rwitajit MAJUMDAR\(^a\), Kensuke TAKII\(^b\), Patrick OCHEJA\(^b\), Mei-Rong Alice CHEN\(^a\), & Hiroaki OGATA\(^c\)

\(^a\)Academic Center for Computing and Media Studies, Kyoto University, Japan  
\(^b\)Graduate School of Informatics, Kyoto University, Japan  
\(^c\)*flanagan.brendanjohn.4n@kyoto-u.ac.jp

Abstract: Digitized learning materials are a core part of modern education and analysis of the use can offer insight into the learning behavior of high and low performing students. The topic of predicting student characteristics has gained a lot of attention in recent years, with applications ranging from affect to performance and at-risk student prediction. In this paper, we examine students reading behavior using a digital textbook system while taking an open ebook test from the perspective of engagement and performance to identify strategies that are used. We create models to predict the performance and engagement of learners before the start of the assessment and extract reading behavior characteristics employed before and after the start of the assessment in a higher education setting. It was found that compared to performance, the prediction of overall engagement has a higher accuracy, and therefore could be more appropriate for identify intervention candidates.

Keywords: Open book; engagement; student performance; reading behavior; online testing

1. Introduction

In the wake of the Covid-19 pandemic, many educators have had to drastically change the way they conduct classes, adapting to online or distance education that has been imposed, and as a consequence to this change in delivery, some traditional methods are difficult to implement. Assessment is often conducted in environments that are tightly controlled including the restriction of reference material. However, it is difficult if not impossible to implement in online and distance education effectively. Open ebook assessments on the other hand can allow the use of reference materials and other sources of information. As assessment is increasingly being performed using LMS and other digital learning systems that are accessed via an internet connection, it is becoming difficult to limit access to external information sources during testing. One argument for open ebook assessment is due to possible information overload of learners. A learner should learn and memorize core knowledge that is key to the domain, while being able to rely on the sourcing of backup or auxiliary knowledge from external references (Heijne-Pennings, 2008). Open ebook testing also can be used to encourage higher cognitive level thinking by reducing the necessity for memorization and rote learning of facts in order to pass a test (Eilertsen & Valdermo, 2000). However, there has been limited work that looks at the actual strategies and information searching through reading behavior that students employ before and during open ebook assessment.

Previously, the authors investigated predicting learner performance from patterns in reading behavior (Flanagan et al., 2020). It was found that reading behavior patterns that occurred during the open book assessment were an important characteristic in predicting low performing learners. In contrast, high performing learners were identified by reading behavior patterns that mainly occurred before the assessment. This indicates the importance of reviewing the lecture materials after class and previewing again before the assessment. While this learning behavior is critical for traditional closed book assessment, previous research suggests that some students view open book assessments differently.
Students might assume that because they will have time during the exam to look at reference materials, prior study is not as crucial to successful performance.

In this paper, we examine the reading behaviors of learners from the first-time topics and concepts are introduced in a lecture, to when they are assessed and the impact of having an open ebook policy assessment. In particular, we are interested in if it is possible to predict problem learners early before the assessment to facilitate possible intervention to improve the learning outcomes. Also, it is beneficial to identify key reading behavior characteristics of high engagement learners when compared to low engagement learners to plan possible effective interventions.

2. Related Work

As the use of digital learning systems is increasing, there are new opportunities to analyze strategies and behaviors of learners from log data that is collected as opposed to more traditional methods of investigation that relied on subjective views and self-reporting from learners. Oi et al. (2015) investigated the preview and review patterns of undergraduate learners by analyzing the usage logs of an ebook reading system. In particular, they examined the aggregate of the number of pages read, the duration of reading and the number of books that were read for a specific time period. It was found that there is a significant difference between the review and preview patterns based on performance in the midterm and term-end examinations.

The problem of predicting low performing students as early as possible has been gaining much attention recently as higher education and MOOCs providers are increasingly examining methods to reduce attrition rates and improve learning outcomes. Okubo et al. (2017) predicted the final academic performance of learners based on their usage of digital learning systems over the course of a 15-week semester. An Recurrent Neural Network (RNN) style model was trained on LMS, e-portfolio and ebook reading events and was able to achieve a high degree of accuracy. Akçapınar et al. (2019; 2019) used features based on aggregates of ebook interaction logs to develop an early-warning system to predict learners that are at-risk of failing the course. 13 different prediction techniques were applied to analyze the data collected over a 14-week semester with promising predictions being made as early as the 3rd week in the semester. Gray & Perkins (2019) demonstrated how at-risk students can be identified within the first 3 weeks by using student attendance/engagement. While Rashid & Asghar (2016) examined the correlation between use of technology/student engagement and academic performance.

Previous research has mainly focused on the analysis of reading behavior for assessment at the semester level or long over the entire span of the course to gain insight to the preview/review patterns or develop early warning prediction models for intervention. In the present research, we focus in particular on the reading behavior of learners in relation to an open ebook assessment. The time frame for intervention in the case that is investigated is also much shorter when compared with previous research and therefore much more fine-grained prediction is required as opposed to prediction at weekly intervals.

3. BookRoll: eBook Reader Log Data

Digitized learning materials are a core part of modern formal education. In addition to serving as a learning material distribution platform, it is also an important source of data for learning analytics into the reading habits of students. The action events of the readers are recorded, such as: turning to the next or previous page, jumping to different pages, memos, comments, bookmarks, and markers indicating parts of the learning materials that are hard to understand or are of importance. The reading behavior of students has previously been used to visualize class preparation and review patterns (Yin et al., 2015; Ogata et al., 2017). The digital textbook system can be used to not only log the actions of students reading reference materials, but also to distribute lecture slides.
In the present work, the non-proprietary BookRoll digital textbook system was used to serve lecture materials and capture learners reading behavior for analysis. As shown in Figure 1, the user interface supports a variety of functions, such as: moving to the next or previous page, jumping to an arbitrary page, marking sections of reading materials in yellow to indicate sections that were not understood, or red for important sections. Memos can also be created at the page level or with a marker to attach it to a specific section of the page. Users can also bookmark pages or use the full text search function to find the information they are looking for later when revising. Currently, learning material content can be uploaded to BookRoll in PDF format, and it supports a wide range of devices, including: notebook computers, tablets, and smartphones, as it can be accessed through a standard web browser.

Reading behavior while using the BookRoll system is send using the xAPI standard for pseudonymized learning event logging and collected in an LRS. Table 1 presents a sample of BookRoll’s learner behavior logs that have been extracted from an LRS. In the logs there are many types of operations, for example, OPEN means that the student opened the e-book file and NEXT means that he or she clicked the next button to move to the subsequent page. An overview of the types of operations and description of the interaction that is represented is shown in Table 2. The logs that are collected in BookRoll are quantitative education data and can be used to observe various objectives, such as (Ogata et al. 2017):

- Analyze the behavior of “active learners” for use in encouraging students to be more active.
- Observe and analyzing the details of behavior of “active learners” to make the students more active.
- Based on the logs made during a class session, improving
course designs, which include collaborative learning and flipped classroom approaches.
- Based on the students’ patterns of viewing e-books (e.g., understanding which page was frequently viewed), improving teaching materials and the structure of the e-books.

Previous work by Authors (Flanagan & Ogata, 2018) details the learning analytics platform that was used to collected the learner behavior data analyzed in the present paper.

Table 1
A sample of events recorded from user interaction with BookRoll.

<table>
<thead>
<tr>
<th>Contents id</th>
<th>Memo text</th>
<th>Operation date</th>
<th>Operation name</th>
<th>Page no</th>
<th>User id</th>
</tr>
</thead>
<tbody>
<tr>
<td>EBOOK 341</td>
<td></td>
<td>2018/01/22 18:10</td>
<td>REGIST CONTENTS</td>
<td>0</td>
<td>t1</td>
</tr>
<tr>
<td>EBOOK 341</td>
<td></td>
<td>2018/01/23 9:16</td>
<td>OPEN</td>
<td>1</td>
<td>s1</td>
</tr>
<tr>
<td>EBOOK 341</td>
<td></td>
<td>2018/01/23 9:20</td>
<td>NEXT</td>
<td>2</td>
<td>s1</td>
</tr>
<tr>
<td>EBOOK 341</td>
<td></td>
<td>2018/01/23 9:21</td>
<td>OPEN</td>
<td>1</td>
<td>s2</td>
</tr>
<tr>
<td>EBOOK 341</td>
<td>Sample memo</td>
<td>2018/01/23 9:22</td>
<td>ADD MEMO</td>
<td>2</td>
<td>s1</td>
</tr>
</tbody>
</table>
Table 2

*Operation names and descriptions for learning behavior interactions captured with BookRoll.*

<table>
<thead>
<tr>
<th>Operation Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPEN</td>
<td>opened the book</td>
</tr>
<tr>
<td>CLOSE</td>
<td>closed the book</td>
</tr>
<tr>
<td>NEXT</td>
<td>went to the next page</td>
</tr>
<tr>
<td>PREV</td>
<td>went to the previous page</td>
</tr>
<tr>
<td>PAGE_JUMP</td>
<td>jumped to a particular page</td>
</tr>
<tr>
<td>ADD_BOOKMARK</td>
<td>added a bookmark to current page</td>
</tr>
<tr>
<td>ADD_MARKER</td>
<td>added a marker to current page</td>
</tr>
<tr>
<td>ADD_MEMO</td>
<td>added a memo to current page</td>
</tr>
<tr>
<td>CHANGE_MEMO</td>
<td>edited an existing memo</td>
</tr>
<tr>
<td>DELETE_BOOKMARK</td>
<td>deleted a bookmark on current page</td>
</tr>
<tr>
<td>DELETE_MARKER</td>
<td>deleted a marker on current page</td>
</tr>
<tr>
<td>DELETE_MEMO</td>
<td>deleted a memo on current page</td>
</tr>
</tbody>
</table>

4. Data Collection and Pre-processing

The data examined in the present paper was collected from an undergraduate Introduction to Informatics course which is a core first year second semester subject at Kyoto University. There were 233 students enrolled in the class. The data was collected for one open ebook assessment that was held at the start of a class for 30min. This was to assess knowledge learnt in a previous lecture. The assessment was provided using the standard testing features on the University’s Sakai LMS. In the weeks leading up to the lecture and assessment, the use of the digital textbook system and testing features in the LMS were introduced and actively used. This ensured that students had good working knowledge of the systems, and were not impaired by using unfamiliar systems. The learning materials for the course were only made available through BookRoll, which has been intentionally designed to restrict offline study by making it difficult to download and print reading materials. The assessment makes up part of the overall final grade of the course. This along with the schedule and focus of the assessments was announced to students at the start and end of each lecture to reinforce the significance of the assessment for learners. As the assessment only focused on one ebook which contained the slides of one lecture, the log data from other learning materials that are not relevant to the assessment were excluded from this study. A total of 164 learners submitted all the questions in the test and were graded.

The lecture which contained the learning material relevant to the assessment was uploaded before being explained to the students. The assessment was given 7 weeks after the lecture, and 5 lectures that focused on a different topic were given during the period until the assessment. The lecture slides were mainly text based with figures and graphs being used sparingly where necessary to assist in explaining models and concepts. The assessment contained questions that involved the simple processing of data or calculation of models which were described in the lecture material that was shared on BookRoll. A short essay was also given at the end of the assessment which asked students to think critically about the possible applications of methods that were introduced during class.

The learning behavior logs were preprocessed to calculate the amount of time a learner spent on each event by comparing the timestamp of neighboring logs. We then removed logs where the learner spent less than 3 seconds on a page as this is indicative of surfing behavior where learners quickly transition from page to page while looking for specific information. Features for training a model were then generated from the filtered raw event logs by concatenating the operation name of four adjacent logs to create a 5-gram feature that represents a sub-segment of the time series of interactions by the learner. This type of feature is also often used in NLP and sequence mining (Flanagan et al., 2014). For example, “NEXT_PREV_NEXT_NEXT_CLOSE” was used when a learner had gone to the next page in the ebook and then returned to the previous page to re-read before reading two the following two pages and finally closing the ebook. In addition, the features were marked with a suffix of “b” or “a” to denote whether the event took place before or after the open ebook assessment had started. The exact time that the learner began the assessment on the LMS was used to account for variations throughout
the class. Learners were divided into two groups based on their performance in the assessment: high and low. The assessment had a maximum score of 17 points, so the groups were divided as follows: low < 8.5 < high. It was confirmed that no learner achieved a score of 8.5 so there weren’t any discrepancies. The groups were nearly balanced, with n=86 for the low group, and n=78 for the high group.

We used the method proposed by Akçapınar et al. (2019; 2019) to measure learner reading engagement from the aggregate frequency of the following reading behavior logs: the total number of reading logs, the number of reading sessions, the total reading time, the number of weeks of reading, the number of days spent reading, the amount of reading events longer and shorter than 3 seconds, how many times the learner went to the next or previous page, number of page jump events, number of red or yellow markers drawn, memo written, and bookmark placed in the learning material. These frequencies were then aggregated using the percentile rank equation as below. It should be noted that the features that were used to calculate the engagement score were not used in the training or evaluation of the prediction models.

\[ PR = \frac{f_b + 0.5f_w}{N} \]

![Graph](image.png)

**Figure 2.** A plot showing the relation between a learners’ score and their level of engagement.

A plot of the learner performance score that was achieved on the open ebook assessment and the overall engagement that was calculated based on the frequency of different types of reading behavior is shown in Figure 2. The correlation between the score that the learner achieved on the open book test and the level of engagement was measured, and it was found that there is a weak correlation of 0.18. This is in contradiction to the results reported by Akçapınar et al. (2019; 2019) and Rashid & Asghar (2016), where it was found that learner performance had a stronger correlation with learner engagement. It can be seen that some students have low overall engagement levels, but have achieved relatively high scores on the test, while other students have high engagement but low scores on the test.

In addition to predicting the score of the student for interventions for at-risk students, the early prediction of a student’s overall engagement could be used to trigger an intervention to increase engagement in the reading task.

The change in learner engagement over time is shown in Figure 3, where at each point in time the engagement of the learner is calculated in relation to their reading behavior up until that point. During the period of the lecture on the left of the x-axis and the open book assessment on the right, there are fluctuations in learner engagement. This could be attributed to learner self-regulation and different learning behavior types, such as: procrastination, learning habit, random, diminished drive, early bird, chevron, and catch-up as described by Goda et al. (2015), and early completers, late completers, early dropouts, and late dropouts as described by Li et al. (2018). Therefore, a learner’s engagement at any
point in time up until the end of the period under examination is not necessarily indicative of the engagement of the learner over the whole period.

Once again, we divided learners into two groups based on their reading engagement: high and low. As the percentile rank is between 0 to 1 the groups were divided as follows: low < 0.5 < high. It was confirmed that no learner achieved an engagement level of of 0.5 so there weren’t any discrepancies, with n=74 for the low group, and n=91 for the high group.

5. Modeling and Analysis of Reading Behavior Characteristics

To model the performance of the learners based on their reading behavior, we approached the analysis as a 2-class classification problem, where the high and low groups were positive and negative class labels respectively. The learners’ raw reading behavior logs were vectorized in the form of the occurrence frequency of 5-gram reading behavior sequence features that are described in the previous section. The vectors were then normalized using the z-score (Kreyszig, 2009), where each dimension in the vector is normalized relative to the sample mean and standard deviation.

First, we examine the problem of early warning prediction. The aim is to identify learners who will have low engagement or performance in the assessment as early as possible before the assessment. The warning could be an intervention that is mediated by the teacher, or an automated intervention, however investigation into this is beyond the scope of this paper and should be addressed in future work. We approach the early prediction problem by training, testing, and evaluating a model for each day between the initial lecture and the start of the assessment where a learner used the digital textbook system to read about the contents of the lecture. Each model was trained on cumulative data up until the day where the prediction is being made, and therefore models are progressively trained on a greater amount of data as the assessment day is closer. At each point, we train a linear-SVM model using weight guided feature selection as proposed by Flanagan, et al. (2014) to select an optimal subset of characteristic features that describe high and low engaged and performing learners’ reading behavior. The performance of the model was evaluated using 5-fold stratified cross validation. These evaluations were then conducted for 30 randomized trials and the average is reported to reduce the possibility of the results being biased due to selective cross validation.

Second, we examine the characteristic reading behaviors of learners from the perspective of high and low engagement and performance before and after the start of the assessment to identify possible differences in strategies that the groups of learners employ for open ebook assessment. In this case, we create an SVM model using all of the available reading behavior data, and add suffixes to identify which events took place before and after the start of the assessment as described in the previous section. Once again, the weight guided feature selection method is used to select a subset of characteristic features of the two groups, and evaluated with 5-fold stratified cross validation over 30 randomized trials. Further,
a test of the feature selection is also conducted to verify the significance of the identified characteristic features.

6. Results

Firstly, we will report the results of the early warning prediction. The evaluation of the SVM model by Area Under Curve (AUC) over time is shown in Figure 4.

![Figure 4](Image)

*Figure 4.* Evaluation of early warning prediction of Engagement and Performance over time by AUC.

The left side of the graph is the day of the lecture where all of the concepts in the assessment are introduced and the learners start reading the lecture materials using the digital textbook system. We can see initially the model cannot predict the engagement or performance of learners with AUC of around 0.5. The first peak in prediction performance is at 30/10 which is the day after the first lecture, with an AUC of 0.7559 for engagement and 0.6405 for performance. Even at this early stage in the prediction, the engagement model is outperforming the performance model by more than 10% AUC. At this point in time the optimal model was trained using only 6 characteristic features for engagement and 80 for performance. The next peak in prediction performance is in the week following the initial lecture with an AUC of 0.7626 for engagement and 0.6475 for performance with 30 and 70 optimal features respectively on around 4-5/11. This could be due to revising by high performing students a week after the initial lecture leading up to the next lecture. The next peak in model performance is the week before the assessment on 4/12, with an AUC of 0.7792 for engagement and 0.6405 for performance with 40 and 30 optimal features respectively. It is possible that this is due to review/preview strategies before the assessment. Finally, the last peak is on 10/12 which is a model trained with all of the data leading up to the assessment that took place on the same day. The final peak was an AUC of 0.8094 for engagement and 0.6499 for performance with both 60 optimal features. It should be noted that the peak in model performance the week after the initial lecture and on the day of the assessment are close, which indicates that in this case predictions and warnings of low performance could be made as early as a week after the initial lecture.

To investigate the strategies that are employed by high and low engagement students, we created a model using all of the available data and tagged the features with a suffix to indicate if the event occurred before or after the assessment had started. A comparison of the 30 trial results for the model using all of the features and the optimized model are shown in Figure 5, where the x-axis is the number of features used to train the model plotted using log scale. The baseline AUC is shown as a dotted horizontal line. We can see that precision initially increases with few features; however, the Accuracy and AUC performance is still low. The model performs best at around 100 optimal features, before declining as additional features are used to train the model.
Figure 5. A comparison of the engagement performance of the baseline model vs feature optimized model.

Figure 6 shows a candlestick plot of the AUC prediction results from the baseline and optimized model. The Shapiro-Wilk test was used to confirm if the 30 trial results for both the original and optimized model are normally distributed (W = 0.9656, p = 0.4267). This indicates that the sample of the trial evaluations had normal distribution in both model results. The Students t-test was employed to determine the significance of the trial results. It was found that the prediction performance measured by AUC of the optimized model was significantly better than that of the original model with p < 0.02.

Figure 6. A box plot of the 30-trial evaluation by AUC.

Table 3 contains the detailed precision, recall, F1, accuracy and AUC evaluation metrics of the optimal performing model. The significance of the F1, accuracy, and AUC were tested using the Students t-test, and all had p < 0.02 indicating that there is a significant difference in the performance of the original and optimal feature model.

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
<th>Accuracy</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>0.9483</td>
<td>0.5652</td>
<td>0.7078</td>
<td>0.7430</td>
<td>0.7634</td>
</tr>
<tr>
<td>Optimized</td>
<td>0.9529</td>
<td>0.8560</td>
<td>0.9017**</td>
<td>0.8971**</td>
<td>0.9018**</td>
</tr>
</tbody>
</table>

*p < 0.05, **p < 0.02
Table 4

*Characteristic Reading Behaviors of High Engagement Learners.*

<table>
<thead>
<tr>
<th>Reading Behavior Sequence</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEXTb CLOSEb OPENb CLOSEb OPENa</td>
<td>0.0742</td>
</tr>
<tr>
<td>OPENa NEXTa NEXTa NEXTa NEXTa</td>
<td>0.0708</td>
</tr>
<tr>
<td>CLOSEa OPENa NEXTa NEXTa NEXTa</td>
<td>0.0637</td>
</tr>
<tr>
<td>NEXTa CLOSEa CLOSEa OPENa NEXTa</td>
<td>0.0616</td>
</tr>
<tr>
<td>CLOSEa CLOSEa OPENa NEXTa NEXTa</td>
<td>0.0569</td>
</tr>
<tr>
<td>NEXTa NEXTa CLOSEa OPENa NEXTa</td>
<td>0.0552</td>
</tr>
<tr>
<td>NEXTa PREVa NEXTa NEXTa NEXTa</td>
<td>0.0543</td>
</tr>
<tr>
<td>CLOSEb OPENb CLOSEb OPENa NEXTa</td>
<td>0.0539</td>
</tr>
<tr>
<td>OPENb CLOSEb OPENa NEXTa NEXTa</td>
<td>0.0539</td>
</tr>
<tr>
<td>NEXTa NEXTa CLOSEa CLOSEa OPENa</td>
<td>0.0535</td>
</tr>
</tbody>
</table>

Finally, we interpreted the features that were used to train the optimal model and the weight that was assigned, which indicates the importance of the feature in predicting high and low engagement learners. The top 10 characteristic reading behaviors of high engagement students is shown in Table 4. It should be pointed out that there are not markedly more features that occur before the start of the assessment as was identify in previous work by Flanagan et al. (2020), and instead all features contain some behaviors that occurred after the assessment started. The characteristic reading behaviors of low engagement students is shown in Table 5.

Table 5

*Characteristic Reading Behaviors of Low Engagement Learners.*

<table>
<thead>
<tr>
<th>Reading Behavior Sequence</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEXTb NEXTa NEXTa NEXTa CLOSEa OPENa</td>
<td>-0.0429</td>
</tr>
<tr>
<td>NEXTa NEXTa NEXTa NEXTa OPENa NEXTa</td>
<td>-0.0330</td>
</tr>
<tr>
<td>NEXTa OPENa NEXTa NEXTa NEXTa NEXTa</td>
<td>-0.0330</td>
</tr>
<tr>
<td>NEXTa OPENa NEXTa NEXTa NEXTa NEXTa</td>
<td>-0.0330</td>
</tr>
<tr>
<td>CLOSEb NEXTa NEXTa CLOSEa NEXTa</td>
<td>-0.0286</td>
</tr>
<tr>
<td>PREVa CLOSEa NEXTa CLOSEa OPENa NEXTa</td>
<td>-0.0262</td>
</tr>
<tr>
<td>OPENb NEXTb CLOSEb NEXTa</td>
<td>-0.0253</td>
</tr>
<tr>
<td>OPENb NEXTb CLOSEb CLOSEb NEXTb</td>
<td>-0.0241</td>
</tr>
<tr>
<td>OPENb NEXTb NEXTa NEXTa NEXTa</td>
<td>-0.0239</td>
</tr>
</tbody>
</table>

7. Discussion and Conclusion

In the present study, we propose and evaluate a method for early warning prediction of high and low engagement students on open ebook assessment. In addition, we also investigate what reading behavior strategies are employed by high and low engagement students. It was found that strategies, such as: revising and previewing are indicators of how a learner will perform and their overall engagement in an open ebook assessment. We anticipate that the use of both early warning prediction of overall engagement and performance could be effective in providing timely interventions to nudge learners to action at key periods.

There are several limitations to the study presented in this paper that should be noted. The number of learners that were observed in this study was restricted to one class, and might be limited for general application. While the features analyzed in this research are not content specific as page numbers and domain information was not part of the feature set, other content level limitations, such as number of pages could impact on the usefulness of the method for other classes or materials.
In future work, we plan to integrate a knowledge map that will provide concept level features of reading behavior to see if it can increase the discrimination of the model and also provide insight into the relationship of assessment items and lecture materials.

Acknowledgements

This work was partly supported by JSPS Grant-in-Aid for Scientific Research (B)20H01722, JSPS Grant-in-Aid for Scientific Research (S)16H06304 and NEDO Special Innovation Program on AI and Big Data 18102059-0 and SPIRITS 2020 of Kyoto University.

References


Applying Learning Analytics to Map Students’ Self-Regulated Learning Tactics in an Academic Writing Course

Ward PEETERS*ab, Mohammed SAQRc & Olga VIBERGd

*Kanda University of International Studies, Japan  
b University of Antwerp, Belgium  
c University of Eastern Finland, Finland  
d KTH Royal Institute of Technology, Sweden  
*peeters-w@kanda.kuis.ac.jp

Abstract: Academic writing is a complex and challenging language learning activity, in which self-regulation is a key critical factor for learner success. Today, a large number of academic writing activities occur in digital learning spaces, including computer-supported collaborative learning settings. Recent advances in the fields of learning analytics (LA) and computer-assisted language learning have provided new opportunities, in part because of the accessibility to new digital learner data, to better understand and ultimately support students’ self-regulated learning (SRL) processes. Even though some related efforts have been made, there is yet a paucity of research targeting foreign language students’ interactive SRL behaviour in online environments. This study aims to shed more light on this issue, and on the possible ways to fill this gap. We used LA methods (frequency analysis, network analysis, statistical analysis and process mining) to analyse and visualise students’ SRL tactics when collaborating with their peers on academic writing tasks on social networking sites. The dataset was obtained from a case study performed at the University of Antwerp (Belgium). In this study, a private Facebook group was integrated in an academic writing course for first-year foreign language majors of English (n=124) and served as an online collaborative space for peer review. The results show, firstly, that foreign language learners use a range of SRL tactics to manage their academic writing process and, secondly, that the strategic SRL task phases (i.e., plan, perform and evaluate) are strongly interconnected. Learners exhibit a readiness and willingness to activate knowledge, monitor progress, interact to adjust to the socio-cultural context and form an identity as a learner. There is a significant positive correlation between students’ use of SRL tactics and their learning performance, which provides novel ground for designing and providing relevant SRL support mechanisms in computer-supported collaborative learning.

Keywords: Academic writing, social networking sites, self-regulated learning, learning analytics, higher education

1. Introduction

In recent years, the fields of computer-assisted language learning (CALL) and self-regulated learning (SRL) have grown closer with an increased interest in the application of data-driven methods to study language learners’ self-regulatory learning processes in online learning settings (Li, Flanagan, Konomi, & Ogata, 2018). Significant attention has been paid to how online spaces can foster, among others, language learner autonomy, self-regulation and a range of productive skills like writing and speaking in second language learning (Montrul, 2019; Peeters & Ludwig, 2017; Viberg, Mavroud, & Ma, 2020). In her review of social networks for language learning beyond the classroom, Zourou (2019) emphasises that Web 2.0 platforms and social networking sites (SNSs) have taken a ubiquitous role in this regard because they have offered educators and researchers new opportunities to foster both formal and informal learning in their daily practice, and have given them access to new types of learner data. The use of SNSs provides an opportunity to balance “the learning benefits from emergent user-driven agency of everyday use with the demands to meet formal curriculum-driven objectives” (Reinhardt, 2017, p.1), and thus forms a sound testing ground for transferable SRL skills in collaborative learning...
settings. This study focuses on academic writing as it is one of the major skills second language learners need to acquire to achieve their learning goals. Academic writing is also one of the skills learners have been observed to struggle most with when entering university, and thus for which they require ample SRL skills (Van de Poel & Gasiorek, 2012). However, in the field of CALL, little research has been conducted on learners’ interactive behaviour in online spaces or SNSs (Peeters, 2019), let alone how interactive behaviour in digital text relates to SRL strategies and tactics for academic writing.

Recently, a growing shift to measure students’ SRL behaviours using student trace log data and multimodal learner data has been kindled by the increasing interest in learning analytics (LA; Viberg, Khalil, & Baars, 2020). The application of LA methods allows scholars to study multifaceted aspects of students’ SRL and, based on these measurements, to offer relevant learner support and optimisation of SRL and the contexts in which it occurs. This study employs LA methods to analyse students’ SRL activities or tactics (i.e., the specific, applied way/s in which a strategy is being used to meet a goal in a certain situation; Oxford, 2016) when collaborating with their peers on academic writing tasks on SNSs in order to reveal how learners use such tactics to optimise their learning. The following research questions have been posed:

1) How do foreign language learners use self-regulatory tactics to manage their academic writing process?
2) How do high and low performing learners use self-regulatory tactics?

2. Background

2.1 Academic Writing and Self-Regulation

Academic writing serves several academic objectives such as promoting critical thinking skills, stimulating creativity, and encouraging discourse as part of a professional community (Hamman, 2005). Consequently, students’ ability to present information and ideas through their writing has “an integral role in academic and professional success” (Applebee et al., 1994, p. 25). Academic writing is a complex and demanding learning activity, in which self-regulation and self-efficacy are critical factors for learner success (Golombek, Klingsieck, & Scharlau, 2019). Self-regulation in writing refers to the “self-initiated thoughts, feelings, and actions that writers use to attain various literary goals, including their writing skills as well as enhancing the quality of the text they create” (Zimmerman & Risemberg, 1997, p.76). When writing in an academic context, learners have to “negotiate rules and mechanisms while maintaining a focus on the overall organisation, form and features, purposes and goals, and audience needs” (Harris et al., 2002, p.110). In such an undertaking, learners employ various SRL strategies and tactics to regulate the complex writing process (Zimmerman & Risemberg, 1997). The importance of fostering SRL strategies and skills among learners, and among foreign language learners in particular, have been stressed by scholars (see e.g., Abadikah, Aliyan, & Talebi, 2018; Csizer & Tanko, 2015). Also, earlier research indicates that there is a correlation between students’ self-regulation of writing and learning performance. Zimmerman and Bandura (1994) for example, found that students’ perceived self-efficacy beliefs in academic achievement and self-regulation of writing could predict their final writing course grade.

2.2 Learning Analytics and Self-Regulated Learning

LA is a growing area of practice and research, focusing on the improvement of student learning outcomes and the improvement of learner support (e.g., support for the development of learner skills such as SRL skills, e.g., goal setting and effort regulation) and teaching (Ferguson & Clow, 2017). One of the emerging subfields that LA researchers target is self-regulated learning (SRL; see Viberg et al., 2020). SRL—which consists of “the processes whereby students activate and sustain cognitions, affects, and behaviors that are systematically oriented toward the attainment of personal goals [i.e., to succeed in academic writing in our study]” (Zimmerman & Schunk, 2011, p.1)—has been found to be critical for second language learners (Oxford, 2016; Viberg & Andersson, 2019; Viberg, Mavroudi, & Ma, 2020) and also for the learner success in academic writing (e.g., Rapp & Ott, 2017).
LA for SRL consists of two constituting parts: 1) a description of student SRL activities, based on traces of actions performed during study (i.e., measurement), and 2) a recommendation, i.e., what should be changed about how learning activities are performed, and instructions about how to change them in order to better achieve learning goals (Winne, 2017). This study mainly focuses on the potential of LA to understand student SRL in a computer-supported collaborative writing setting. The findings of the analysis aim to provide a sound ground for relevant recommendations to improve learner SRL support in the chosen learning settings. Earlier LA research on writing analytics have shown promise of how LA can support academic writing. Notable is the work on the Academic Writing Analytics (AWA) web app, the results of which showed how the provision of actionable feedback helped students improve their writing skills (Gibson et al., 2017). Similarly, within the frames of the same AWA project, an LA tool (AcaWriter) offered students personalised feedback on their writing (Knight et al., 2020). Others have also shown the value of using automatic detection of meta-discourse to support academic writing (Bektih, 2017). However, research on self-regulation and academic writing remains scarce and this study aims to further inquire into this issue to fill this gap.

2.3 Theoretical Lens of Self-Regulated Learning

To understand students’ SRL tactics in computer-supported collaborative writing settings, and in particular, to categorise them according to the SRL phases, the theoretical lens of the Strategic Self-Regulation (S2R) Model of language learning (Oxford, 2016) has been adopted. In general, the S2R model targets meta-strategies, strategies and tactics that language learners can use to regulate various aspects of their learning, including beliefs, observable behaviours, their internal mental states, as well as the learning environment. The model includes a sequence of phases for doing a task (Figure 1). In task-phase 1, strategic forethought, the learner pays attention to the demands of the task, goals, plans how to address them, and activates them. In the second phase, strategic performance, the learner implements the plan, monitors how it is working, and decides whether to continue the task as it is going, stop, or make changes in how to approach the task. In the third phase, strategic reflection and evaluation, the learner makes judgements of value about learning outcomes, the effectiveness of selected strategies and tactics, and the self (e.g., self-efficacy).

![Figure 1. Strategic self-regulation task-phase model of language learning (Oxford, 2016).](image)

As mentioned above, the S2R model includes not only meta-strategies and strategies, but also tactics, which are focused on in this study. Tactics refer to more specific manifestations of a strategy and meta-strategy by a particular learner in a given learning setting for a specific purpose. They denote the specific, applied way/s in which a strategy is being used to meet a goal in a certain situation.

2.4 Case Study Settings

The dataset used in this paper was obtained from a case study performed at the University of Antwerp (Belgium). In this study, a private Facebook group was integrated in an academic writing course for first-year foreign language majors of English (n=124) and served as an online collaboration space for peer review in which they could share their written work, discuss their progress and ask questions about their writing and learning process. The course consisted of 12 contact hours, blended with an online self-access module with exercises on academic literacy, and the peer review forum on Facebook. No
teachers were included in the Facebook group. Learners had to rely on their peers for support (Murray, 2014). Over a period of three months, learners had to write three 300-word essays. After an initial brainstorm in class, learners had to finish their essays at home and were informed that they could consult with their peers on Facebook about their writing process and the challenges they faced.

3. Methods

In the initial stages, the data set was manually coded using the principles of digital conversation analysis to distinguish recurring patterns in the structure and content of status updates and comments (cf. Farina, 2018). In several data-driven coding phases (cf. DeCuir-Gunby, Marshall, & McCulloch, 2011), a team of two coders compiled a taxonomy of core activities in the data sets (Peeters, 2018), after which a team of four coders checked the coded transcripts. The team reconciled to discuss coding errors and inter-rater reliability. Disputed codes were further discussed and amended until a consensus was reached on all the codes in the transcripts. The taxonomy presented below (Table 1) was used as the basis for this study into students’ SRL activities in an academic writing course. The table includes the topics and tactics students addressed (adapted from Oxford’s (2016) cognitive, affective and sociocultural-interactive activities), representing the task-phases in the S2R model (Figure 1).

| Students sharing stories, tips and tricks about academic, cultural, social, psychological and linguistic challenges they face. | Acculturating |
| Students planning the next steps in their writing or learning trajectory and implementing those plans. | Planning |
| Students discussing and familiarising themselves with goals, objectives and requirements of the course and tasks. | Organising |
| Students discussing vocabulary, jargon, grammar and textual structure while writing essays. | Writing text |
| Students discussing topics / thesis statements for their essays, and discussing reasoning and logic of their (counter-)arguments. | Argumenting |
| Students sharing, discussing and evaluating resources provided by the university and by the peer group. | Using resources |
| Students talking about hobbies, free time and leisure. | Social bonding |
| Students expressing positive feelings towards their peers, acknowledging their work and thanking them. | Acknowledging |
| Students discussing and applying feedback they received from the teacher or from their peers. | Feedback |
| Students discussing the purpose and organisation of the course, the tasks and the collaboration between peers. | Reflecting |

Table 1. Overview of Tactics and S2R task phases in the data set

Learners received grades through continuous assessment throughout the semester. The essay-writing task added up to 30% of the overall grade for the course, next to in-class tests and assignments, as well as self-access exercises online. Learners’ overall grades are compared to their interactive behaviour online to study if high-performing learners use different SRL tactics compared to low-performing learners in academic writing. Previous research in process mining and analysis of SRL strategies has shown that the comparison of deciles helps show the differences or similarities between subgroups in how they implement learning strategies or tactics (Saint, Gašević, Matcha, Uzir, & Pardo, 2020; Saqr et al., 2019). We accessed the scores on the three essay-writing tasks and compiled a composite score. The top decile was identified as the top 10 percent performing students, the bottom decile was identified as the bottom 10 percent performing students.
3.1 Data Extraction

A Graph API Explorer tool was used to scrape and compile the dataset. The dataset, consisting of 2594 entries, includes IDs for every participant, status updates and comments, post IDs for every entry and time stamps. The data set has been anonymised before any analysis was conducted. Since some posts were coded with two tactics, the final count was 3123.

3.2 Analysis

To answer the research questions and reveal the multifaceted nature of self-regulated academic writing, multiple analytical lenses were used. We implemented four analytical methods: 1) frequency analysis, 2) network analysis, 3) statistical analysis, and 4) process mining. They have been employed to highlight different aspects of students’ tactics and their correlation with students’ performance. For RQ1 the four methods were used to reveal different tactics and how they correlate to each other. For RQ2, we present a comparison between the high and low performers. Frequency analysis was conducted to demonstrate the distribution of tactics among learners (RQ1) and to report the differences and similarities between learners’ subgroups (RQ2). Frequency analysis was conducted with the R programming language. Statistical analysis was performed to investigate the correlation between the employed tactics and students’ performance (RQ2), with the aim to highlight frequent tactics that could serve as indicators of favourable course outcomes. The correlation was performed using R and Spearman's rank correlation coefficient since the grades were not normally distributed, as examined by the Kolmogorov-Smirnov and the Shapiro-Wilk tests of normality. Process mining was implemented for determining and visualising the processes of the students’ interactive behaviour and tactics.

Two types of process mining approaches were applied. Firstly, relative frequency-based process mining offered by the R package BupaR was used (Janssenswillen, Depaire, Swennen, Jans, & Vanhoof, 2019). Bupar packages offer sequential process maps that highlight the flow and frequencies of examined tactics. To construct the process maps, the timestamp of each student interaction was used as the event time; the coded tactic was used as the ‘event’ and the students’ IDs as the case IDs. The node metrics of the process map represent the relative frequency of the implemented tactic; the edges represent the associative internode relative frequencies (RQ1). Secondly, stochastic process mining was performed using the R package PMineR to explore the associative probabilities of students’ tactics (Gatta et al., 2017). PMineR offers methods for the calculation and visualisation of First Order Markov Models (FOMMs). FOMM computes the probability that a current learner tactic depends on the previously employed tactic known as the transition probability (TP). TP is reflected in the FOMM models through the width of the edges between the FOMM nodes. The FOMM graphs were generated for the whole learner group (RQ1), and to compare the top vs the bottom deciles (RQ2). A directed network was constructed by considering the reply of a post as source, and the replied to post as target (RQ1). For each tactic, we calculated degree centrality as the number of interactions for the tactic (incoming and outgoing), betweenness centrality as the number of times a tactic has lied on the shortest path between others, thus connecting them; closeness centrality as the inverse distance between a tactic and all others, diffusion centrality as the probability that tactics propagate/stimulate further interactions.

4. Results

The frequency of the tactics used by learners (Figure 2) shows that posts and comments that addressed issues with writing text (16.8%) and constructing argumentation (16.1%) for essays took up a considerable part of the peer review process. Given the nature of the assignments and the educational context, this was expected. Providing acknowledgement (20.6%), which includes expressing positive feelings towards peers, acknowledging their work and thanking them, was also very prominent. These three tactics are part of the strategic performance phase in the S2R model (Oxford, 2016). Other tactics in this phase, in decreasing order of frequency, could be found in posts and comments that revolved around social bonding (8.3%), applying feedback (7.2%) and using resources (6.1%). Acculturating to the academic context (9.9%), which concerns discussions between peers about the challenging transition from their previous scholarly experiences towards a university’s academic culture (Peeters &
Fourie, 2016), was the most prominent tactic when it comes to strategic forethought. Tactics in which students familiarise themselves with the goals, objectives and requirements of the course and tasks (9.4%) came in as a close second. Planning (0.4%), on the other hand, was a tactic seldom used by students in their conversations with others. Within strategic reflection and evaluation, students reflected on their performance (5.3%), yet less frequently than other tactics. Students predominantly engaged in the performance phase while also paying attention to the demands of the task, set goals, plan how to address them, and activate existing knowledge.

Operationalising the interactions as a network of tactics shows how interactions were built, and which tactics were well-connected or central (Figure 2). Expectedly, writing text and arguments were the most prominent central tactics in terms of quantity (Table 2). However, acculturating and organising were the tactics that were most likely to connect to other tactics (high betweenness centrality degree). Acculturating, acknowledging, reflecting and social bonding were the tactics close to all other tactics. Using diffusion metrics to discover which tactics were more likely to propagate or stimulate further discussions, we see that almost all tactics were well connected, but that acculturating, acknowledging, reflecting and social bonding, again, were the tactics that were the most central. Notably interesting are the tactics with a 0 score for betweenness centrality: acknowledging, applying feedback and writing arguments, indicating that these tactics were less connected to other tactics or were predominantly used to conclude or end a conversation.

Figure 2. Summary of tactics in the peer interaction process on academic writing (left) and network of students’ SRL tactics (right). The nodes represent different tactics while the directed edges represent the degree of connection between them.

Table 2. Centrality Degree Scores of Students’ SRL Tactics

<table>
<thead>
<tr>
<th>Tactics</th>
<th>Degree</th>
<th>Closeness</th>
<th>Betweenness</th>
<th>Diffusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acculturating</td>
<td>0.46</td>
<td>1.00</td>
<td>0.13</td>
<td>1.00</td>
</tr>
<tr>
<td>Acknowledging</td>
<td>0.37</td>
<td>1.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Applying feedback</td>
<td>0.10</td>
<td>0.49</td>
<td>0.00</td>
<td>0.96</td>
</tr>
<tr>
<td>Organising</td>
<td>0.63</td>
<td>0.72</td>
<td>0.12</td>
<td>0.99</td>
</tr>
<tr>
<td>Planning</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>Reflecting</td>
<td>0.22</td>
<td>1.00</td>
<td>0.08</td>
<td>1.00</td>
</tr>
<tr>
<td>Social bonding</td>
<td>0.28</td>
<td>1.00</td>
<td>0.01</td>
<td>1.00</td>
</tr>
<tr>
<td>Using resources</td>
<td>0.07</td>
<td>0.49</td>
<td>0.10</td>
<td>0.95</td>
</tr>
<tr>
<td>Writing arguments</td>
<td>1.00</td>
<td>0.72</td>
<td>0.00</td>
<td>0.99</td>
</tr>
<tr>
<td>Writing text</td>
<td>0.94</td>
<td>0.72</td>
<td>0.05</td>
<td>0.99</td>
</tr>
</tbody>
</table>

A correlation matrix was generated between tactics and students’ performance (Figure 3). There was a positive and significant correlation between the assignment grades and the frequency of 1. acculturating ($r_s=0.27$, $p<.01$), 2. acknowledging ($r_s=0.19$, $p<.05$), 3. organising, ($r_s=0.3$, $p<.01$), 4. writing arguments ($r_s=0.22$, $p<.05$), 5. writing text ($r_s=0.38$, $p<.01$), and the total number of interactions with peers regardless of the type ($r_s=0.29$, $p<.01$). The correlation coefficient results also show which tactics were correlated. The highest correlated pairs were: 1. organising and acknowledging ($r_s=0.6$, $p<.01$), 2. organising and acculturating ($r_s=0.56$, $p<.01$), 3. reflecting and social bonding ($r_s=0.55$, $p<.01$), 4.
acculturating and writing arguments \((r_s=0.54, p<.01)\) as well as 5. reflecting and acculturating \((r_s=0.53, p<.01)\). These results also stress the importance of the tactics of writing text, argumentation and acculturating as indicators of success in academic writing, as well as in kindling other successful tactics.

**Figure 3.** Correlation matrix between students’ use of tactics and students’ performance.

A closer look at the sequential process map of students’ SRL may explain the sequential process of the discussions (Figure 4). Students are likely to start with composing argumentation for essays (54.5%), composing text (29.4%) or sharing questions and thoughts on the academic context they have to function in (10.1%). Acknowledging is the step that most frequently follows other steps in this process, often followed by other acknowledgements (54.3%). Similarly, steps that involve writing arguments are often followed by steps that address argumentation (45.2%). Writing text also has this feedback loop (40.4%). The frequency-based graphical mapping shows a general map of the process. Yet, it is limited in making inferences about the likelihood of transitions. Hence, FOMM graphs are necessary.

**Figure 4.** Sequential process map of students’ SRL tactics

The FOMM graphs (Figure 5) show a more diverse range of starting points for sessions of interactions, with the highest likelihood of starting with writing text \((0.37)\), followed by writing arguments \((0.2)\), or organising and acculturating with both a 0.12 probability. Acknowledging is a common transition step to follow other steps with higher probability than other tactics. Similarly, writing text acts as a common

---

**Figure 5.** The FOMM graphs showing the diversity of starting points for sessions of interactions.
transition step after reflection, writing arguments, organising and acculturation. This mapping reflects a diverse approach to SRL with a set of prominent steps that seem to be central in the interaction process.

In order to answer the second research question and analyse whether the top and bottom decile act differently, FOMM graphs for the two groups were created (Figure 6). The results show that top achievers are more likely to initiate a conversation by using the tactic of writing text (0.5) compared to bottom achievers (0.31). Bottom achievers also are more likely to use a range of different tactics to start a conversation, including using resources (0.23), acculturating (0.23) and organising (0.13) in addition to writing. Top achievers are more likely to address issues with writing arguments (0.17) and organising (0.17) at the beginning of a conversation. Both groups are likely to acknowledge others as a follow-up step in the interaction process.

**Figure 6:** First Order Markov Model of top and bottom achievers’ SRL tactics

5. **Discussion & Conclusions**

This study aimed to analyse how foreign language learners’ interactive behaviour in digital text relate to their SRL tactics in an academic writing course. The findings of this study exhibit that foreign language learners use a range of SRL tactics to manage their academic writing process. Composing text and argumentation—which represent the strategic performance phase of the S2R model (Oxford, 2016)—are frequently used tactics and from the results of our sequential process mapping, we found that these tactics are often used as conversation starters. Correlation coefficient findings also show that certain tactics are highly correlated. Most notably, correlations between organising (a part of the strategic forethought phase) and acknowledging (the strategic performance phase) indicate that students often express positive feelings towards each other in their discussions on how to deal with the course, its tasks
and its requirements. Organising and acculturating are also linked, which suggests that discussing goals, objectives and requirements (the strategic forethought phase) often go hand in hand with adjusting to the broader sociocultural context of academia. Reflecting and social bonding correlate as well. This exhibit that reflections might often coincide with telling personal stories and self-disclosure. In addition, the network analysis has shown the prominence and interconnectivity of tactics like acculturating, acknowledging, reflecting and social bonding within a wide range of interactions. These findings demonstrate to what extent the S2R task phases are interconnected, and how learners can orchestrate them, with room for activating knowledge, monitoring progress, interacting to adjust to the socio-cultural context and forming an identity as a learner.

To answer the second research question, we analysed if high-performing learners used different SRL tactics compared to low-performing ones. The results have shown that students who paid special attention to the form and/or content of their writing, as well as the ones who acknowledged others during discussions, scored higher marks than those who did not. Those who engaged in the strategic forethought phase of academic acculturation and organisation, i.e., negotiation context (cf. Harris et al., 2002), also performed significantly better. Similarly, students who discussed how to apply feedback from peers or tutors were more likely to score higher final course grades.

A positive and statistically significant correlation was found between the final course grades and writing text, writing arguments, and discussing goals, objectives and requirements of the course and tasks. The FOMMs comparing high and low achieving students echo these findings and additionally show how the top decile is more likely to initiate conversation using tactics from the strategic performance phase: writing text and arguments. The bottom decile uses a wider range of tactics like using resources, acculturating and organising as well as writing text when starting a conversation. While both groups make use of tactics that represent strategic forethought, the top decile is more likely to dive head first in the actual writing process when interacting with their peers. It has to be noted that both groups rarely discuss planning, and that the top decile even tends to skip this tactic altogether.

In summary, this study has offered a solid example of how students’ SRL tactics can be effectively traced with the help of LA methods. To obtain a more holistic picture of the students’ SRL process, the next step will be to examine how these tactics are linked to other constructs (i.e., SRL strategies and meta-strategies) and dimensions (i.e., cognitive, affective and socio-cultural interactive) of the used S2R model (Oxford, 2016). This will offer a more sound ground for designing and providing relevant SRL support mechanisms, for example in a form of student-facing learning dashboards (Viberg et al., 2020). Such dashboards can also be oriented towards teachers for the purpose of enabling them to teach relevant SRL strategies and intervene in time.

References


Experimental Design of Automated Extraction for 3-Level Tutoring Support Information in Programming Exercises

Yasuhiro NOGUCHI*, Kousei AYABEa, Koichi YAMASHITAb, Satoru KOGUREa, Raiya YAMAMOTOC, Tatsuhiru KONISHId & Yukihiro ITOHd

a Faculty of Informatics, Shizuoka University, Japan
b Faculty of Business Administration, Tokoha University, Japan
c Faculty of Engineering, Sanyo-Onoda City University, Japan
d Shizuoka University, Japan
*noguchi@inf.shizuoka.ac.jp

Abstract: In programming classes, teachers and teaching assistants (TAs) cooperate to support learners’ programming exercises. In many situations while providing support, however, teachers cannot find and resolve the impasse for individual learners while at the same time monitoring the impasse trend in the entire class to effectively provide learners with additional information. In this paper, we categorized 3-level tutoring support information for teachers’ activities and designed the experimental architecture for a system that supports automated extraction for the 3-level information from learners in programming exercise activities.

Keywords: Automatic impasse detection, programming exercise, monitoring, decision support

1. Introduction

In recent years, the development of an information society has required more information engineers, including system engineers. In this context, programming skills are not only necessary for such engineers, but are becoming fundamental skills for most of them to utilize information systems in their fields. As well as science or engineering colleges and technical schools for such engineers, liberal arts colleges and ordinary high schools also provide some programming education classes. In early programming education, the following two teaching methods and their mixed method like PBL are often used to acquire basic programming skills:

- A lecture-style method whereby students learn the theory of programming, syntax of programming languages and behaviors, and features of algorithms by listening to the teacher’s lecture and reading textbooks themselves.
- An exercise-style method whereby students learn the practical skills of programming by completing tasks based on the lecture content. While the learners attempt to complete the exercises, the teachers circulate among the learners to answer their questions, check the work of learners who are having difficulties, and to provide learning support.

However, novice learners occasionally experience impasses during coding exercises. They might be unable to recover from an impasse for a long time, and therefore cannot proceed with their exercise. For example, syntax errors are the most common errors for them; as reported by Denny, Luxton-Reilly, David, and Hendrickx (2011) based on their analysis of Java programing exercise classes, “some students were often unable to solve their syntax problems.” Becker et al. (2011) summarized various viewpoints for improving compilers’ error messages for novice programmers (readability, cognitive load, provide context, show examples, scaffolding, logical argumentation, and so on). It has been shown that various
measures are necessary to convey the message to learners. Furthermore, learners often reach an impasse where they cannot construct the steps for implementation, cannot find the location (line number) of the codes that have caused the runtime error, cannot solve the run-time error, cannot complete the procedures to attain the expected output values with the expected behavior of the program, and other problems. The cause of an impasse basically depends on the individual situation. Moreover, learners cannot identify the impasse situation, and even if they do have opportunities to ask teachers, they may not be able to explain the situation correctly.

In programming education, many learners complete the exercises using their own PCs, while a small number of teachers and teaching assistants (hereinafter, “teachers”) support them by circulating to answer their questions. Recent novice programming classes are typically one-to-many teaching situations that may face the following difficulties:

A) Difficulty in finding and resolving the impasse for individual learners
   Since the learner completes the exercise using their own PC, it is difficult for the teachers to identify what student has reached an impasse in the exercise. Moreover, when the teachers take the learner’s PC to identify the cause of their impasse, it shortens the time the learner has in which to complete the exercises.

B) Difficulty in monitoring the overall class trend in learners’ impasses
   Since individual teachers do not have time to share their support results, it is difficult for them in the one-to-many teaching situation to monitor the overall impasse trend in the class in real time. Thus, in the class, it is difficult for teachers to provide additional instructions to learners based on the common types of impasse and common causes of impasse amongst the class learners. Moreover, they cannot update their teaching materials based on the types of impasse identified.

Regarding the detection of impasses and mistakes, several methods have been proposed to detect these by collecting and analyzing learners’ compilation errors and by analyzing learners’ programming activities such as keyboard, mouse click, et cetera (Mazza & Dimitrova, 2004; Biswas & Sulcer, 2010; Hartmann, MacDougall, Brandt, & Klemmer, 2010). However, although some automated feedback tools focus on error detection, they do not provide sufficient feedback for novice programmers (Keuning, Jeuring, & Heeren, 2008). McCall and Kolling (2019) show that there is a large difference in the frequency of occurrence and difficulty of each type of problem for novice programming learners. Therefore, it is difficult to provide guidelines in advance for the problems that may occur. Furthermore, Brown and Altadmiri’s (2014) study showed some discrepancy between the frequency of problems recognized by teachers (like many learners being troubled by a particular problem) and the actual frequency of learners’ problems. It is necessary, therefore, to provide individualized tutoring based on the actual learner’s situation, and to rely less on teachers’ intuition. It is also important to monitor the real-time learners’ impasse trend in the classroom.

To reduce these difficulties, in this research, we classified information supporting the difficulties A) and B) into three levels that can be analyzed with the impasse detector (Yamashita et al., 2017). We suggested an experimental design for the system supporting automated extraction of the 3-level information from learners.

2. 3-Level Tutoring Support Information

In this section, we describe the requirements for the monitoring system to support the role of teachers in the exercises portion of learning programming: circulating among the learners to answer their questions, providing additional instructions for them, and updating teaching materials to eliminate their impasses. To reduce the difficulties A) and B), the following functions are required for the system:
A1) A function that provides teachers with a trigger for circulating to a specific learner by automatically identifying whether the learner is in an impasse or not.

A2) A function that provides teachers with tutoring information regarding the learner who has reached an impasse without suspending the learner’s programming activities before the teacher circulates to that learner.

- What is the location of the latest learner’s code in their impasse?
- What is the location of the teacher’s correct answer code corresponding to the latest learner’s code in their impasse?
- What are the causes of the impasse?

B1) A function that identifies the number and tendency for learners to reach an impasse as a trigger for teachers to provide additional instructions and/or to re-explain the exercise to the entire classroom.

B2) A function that extracts common impasses for learners.

- What are the common locations in the teacher’s correct answer code in learners’ impasses in the classroom?
- What are the common causes for learners’ impasses in the classroom and/or the course of the programming exercise class?

In previous research (Yamashita et al., 2017), the impasse detector focused on detecting some signs of impasse based on individual learners’ coding activities in real-time. Teusner, Hille, and Staubitz (2018) also detected some learners’ impasses from the tendency of keyboard input (keyboard interventions, and so on). Therefore, these studies could be positioned to contribute to the realization of the function A1). However, to realize the function A2), the system should identify the location (line number) in the learner’s code at which the learner has reached an impasse. Additionally, the learner’s impasse location in their code, identified in previous monitoring time, should be traced to the learner’s latest code, if the location is moved because the learner has edited the code for another purpose. Moreover, the location should correspond to the location of the teacher’s correct answer code in order to understand the learner’s code in the context of the current programming exercise problems. Furthermore, providing some candidates regarding why the learner has reached an impasse is useful for teachers when preparing their support plan for learners.

Regarding B), it is difficult for a small number of teachers to aggregate individual impasse conditions for many learners. Thus, the monitoring system should support teachers by summarizing and extracting the situations that occur in the classroom. To realize the function B1), not only should the system identify the occurrence of a learner’s impasse, but it should also identify whether the sign of impasse is continuing or not, and when the sign disappears. To realize the function B2), the system should aggregate learners’ codes, and analyze them for correspondences based on the teacher’s model code since each learner’s code is different. If the system aggregates learners’ common impasses in the classroom on the teacher’s model code, the teacher can then prepare additional instructions and identify points of update for the teaching materials on that basis. Additionally, the common causes of learners’ impasses could be aggregated not only as a class, but also whole classes in the course can be used to update teaching materials and scaffolding strategies.

Based on the above, we propose a framework of automated extraction for 3-level tutoring information as shown in Figure 1. Lv.1 is the impasse line number of the location on each learner’s latest code. Lv.2 is the impasse line number of the location on the teacher’s model code corresponding to each learner’s impasse. Furthermore, Lv.2 information shows the impasses common to multiple learners. Lv.3 comprises the learning items that allow learners to resolve their impasses corresponding to each learner’s impasse. The Lv.3 information is aggregated to identify learning items that are common to learners in multiple classes.
3. Experimental Design for Proposed System

3.1 Architecture of the System

Figure 2 shows the architecture of the system. The Learner’s Activity Collector collects the learner’s codes, compiles a log, and provides the execution log of the successfully compiled program whenever the learner compiles their codes. These data are then sent to Impasse Analyzer Lv.1, Lv.2, Lv.3 in sequence, and these analyzers extract Lv.1, Lv.2, Lv.3 information. The teacher creates some model codes (hereinafter, “standard algorithm”) in advance and adds links on certain lines of the model codes to learning items in the learning materials used in the programming course.

3.2 Impasse Analyzer Lv.1

While the existing impasse detector (Yamashita et al., 2017) only identifies whether the learner has reached an impasse or not, our impasse detector observes learners’ coding activities and can observe a learner’s focused location by presenting a sign of the learner’s impasse based on 10 rules in the impasse detector. Thus, for the prototype implementations for the system, a line number is provided when the sign of impasse appears. Additionally, by analyzing the changes from the immediately preceding codes each time, the record in the latest learner’s code succeeds the impasse history at the corresponding location. Finally, the analyzer recodes the impasse types and their accumulated number on the line in the latest learner’s codes, which enables them to continuously measure learners’ impasses at the same location.

3.3 Impasse Analyzer Lv.2

The analyzer records impasse types and their accumulated number and the number of learners who reach the same impasse on the line in the standard algorithm. This is realized by analyzing learners’ codes and applying a standard algorithm to synchronize corresponding lines based on
their code design structures. In many studies, static analysis for the program code is used to create a correspondence between each learner’s codes and the teacher’s model codes. It is known that if the constraints on the learners’ developing codes are strict in the exercise portion of the class, static analysis can be performed for line-by-line correspondence between the learner’s code and teacher’s model code (Keuing, Jeuring & Heeren 2018). However, if the constraints of the learners’ code are strict, this limits the possible exercises that teachers can design, particularly for developing learners’ practical skills.

---

**Figure 2. Architecture of the System (analysis flow example for one learner)**

**Figure 3. Comparison between learners’ codes and standard algorithm in Impasse analysis Lv.2**

Therefore, in this paper, we used the format of “standard algorithm” in Konishi, Suzuki, and Itoh (2000) to allow some variations in the teacher’s model code. The standard algorithm is described in extended PAD (PAD is a structured diagram developed by (Futamura, Kawai, Horikoshi & Tsutsumi (1980)), and enables teachers to write model codes that include variations in the ordered blocks and selective blocks in Figure 3. Most implementations for algorithms in programming classes for novice learners include free ordering blocks (where the same results are executed from proceeding block A and B as in proceeding B and A), and free selective blocks (where the same results are executed from proceeding A and proceeding B). By this means, multiple patterns of model code variations can be prepared relatively easily.
3.4 Impasse Analyzer Lv.3

When teachers create a standard algorithm, they add links from lines on the standard algorithm to learning items to solve the impasse in the lines. The lines in the standard algorithm record the number of learners who have reached an impasse and the number of times an impasse occurs on a given line. As a result, from the findings regarding which learning items are related with a greater number of impasses, teachers can learn which learning items are likely to bring learners to an impasse in the classroom due to lack of understanding.

4. Conclusion

In this research, to reduce teachers’ difficulties when circulating amongst learners during programming exercises, we designed a framework for collecting 3-level tutoring information from learners completing the exercise. We proposed a prototype system to extract this information based on an existing learners’ impasse detector and static analysis between learners’ codes and teachers’ model codes. In the future works, to increase the performance of the function for problem A) and B), it is necessary to tune the threshold function and clarify the restrictions/expectations for learners while adjusting to real programming exercise situations.

Acknowledgements

This study was supported by Japanese Grants-in-Aid for C19K122650.

References


A Prototype Framework for a Distributed Lifelong Learner Model

Patrick OCHEJA a*, Brendan FLANAGANb, Solomon Sunday OYELERE c, Louis LECAILLIEZ a & Hiroaki OGATA b

a Graduate School of Informatics, Kyoto University, Japan
b Academic Center for Media and Computing Studies, Kyoto University, Japan
c School of Computing, University of Eastern Finland, Finland
* ocheja.ileanwa.65s@st.kyoto-u.ac.jp

Abstract: Learning is ubiquitous and as learners change environment, it becomes necessary to collect their learning footprints across multiple systems. However, as these learning footprints are collected as logs on different systems, protected by varying privacy definitions, there exist of problem of not being able to process these logs together in a useful manner. In this work, we build on a blockchain of learning logs platform (BOLL) to construct a learner’s user model based on logs from multiple systems in a distributed method using a decentralized network. We propose a framework that can also connect a learner’s user model from multiple systems into one representative model that can answer questions about the learner’s lifelong learning. Finally, we show typical scenarios of how such lifelong learner models can be used to support teaching and learning.

Keywords: lifelong learning, learning logs, user model, learner model, blockchain, BOLL.

1. Introduction

The use of learning systems has become prevalent in schools and continues to result in more data being logged from students’ interactions. By analyzing these learning logs, it has also become possible to support a more data-driven and personalized education (MacMillan & Schumacher, 2010). This includes supporting at-risk students (Akcapinar et al., 2019), learning content design (Demchenko, Gruengard & Klous, 2014), facilitating group activities (Messina et al., 2013), recommending useful learning contents (Tarus, Niu & Yousif, 2017) and sharing successful learning habits (Zhu, 2012). Although students may attend more than one institution, current learning analytics processes rely heavily on data from students’ current institution regardless of their past learning experiences or learning logs. Thus, the following problems arise:

- It is difficult to fully assess and understand the learner’s knowledge or lifelong learning.
- Inability to personalize learning with minimal effort (the cold-start problem).
- Unable to diagnose and provide concrete answers to a learning difficulty.
- Limitations with sharing successful learning habits at scale.

These problems could be caused by factors such as: the difference in learning tools and data standards, lack of interoperability, the difficulty in facilitating communication between systems due to privacy limitations as well as other ethical concerns.

To solve these problems, we propose a framework that can connect a learner’s lifelong learning across different institutes and learning platforms into a learner model. We acknowledge that different learning systems may store their learners’ data in varying formats. In this work, we focus on exploiting how logs from multiple learning systems with similar data formats can be unified. We demonstrate how learners can also manage their privacy, reflect on their lifelong learning, share and access successful learning habits of other students at different schools in a privacy preserving manner. We also show how teachers and content designers can use the resulting lifelong learning user model to diagnose and solve learning difficulties, and design personalized learning contents.
2. Related Work

There are previous researches on ontology mapping as reviewed by Choi, Song and Han (2006) as well as the use of knowledge or concept maps in learning design (Baker & Inventado, 2014). However, the use of isolated data storage sites and the lack of decentralized technologies (before now) to manage distributed access has made it difficult to connect and construct concept maps from students' lifelong learning. To solve the limitations of disconnected lifelong learning such as the cold-start problem and lack of transferability of learning logs, Ocheja, Flanagan & Ogata (2018) and Ocheja, Flanagan, Ueda & Ogata (2019) proposed and implemented a Blockchain of Learning Logs (BOLL). BOLL is a platform that allows the transfer of learning records through the blockchain and provides a means to manage privacy permissions using smart contracts. However, using these records to realize a user model that can answer some useful questions about a student’s lifelong learning is yet to be considered.

Kay & Kummerfeld (2019) proposed a conceptual model for evaluating how learning applications and data repositories can be used to realize Personal User Model for Lifelong, Life-wide Learners (PUMLs). Our work takes the discussion further by presenting a concrete framework with an existing privacy-first (Conoscenti, Vetro & De Martin, 2017), decentralized application that truly connects lifelong, life-wide learning: a limitation acknowledged by Kay & Kummerfeld (2019). We extend the BOLL platform to enable computing of user model from distributed records and provide a query enabled interface for supporting and personalizing learning.

Platforms like ALEKS, a web-based intelligent tutoring system (Canfield, 2001), have been found useful in diagnosing learning difficulties and supporting learning as demonstrated by Hagerty & Smith (2005). However, the ALEKS platform provides learning assistance by using information within the learners current learning environment. In contrast, our proposed framework diagnoses learning using a learner’s records from various learning environments. It is important to be able to use data from multiple learning environments as such data may provide adequate information to achieve personalized learning.

3. Research Method

The design of the proposed framework is based on participatory design (Simonsen & Robertson, 2012). Participatory design is a co-design approach which involves the active engagement of all stakeholders in the design process in order to ensure the outcome fulfill the needs of the stakeholders. To solve the problems with enabling a connected lifelong learner model, we identified three main stakeholders: students, teachers and administrators. We carried out the co-design process with 4 students, 3 teachers and 3 administrators. Specifically, we discussed with the stakeholders on the current challenges of constructing lifelong learner models from learning logs, identified key factors to solving the problem and tested various solutions.

4. Proposed Framework

We propose a privacy-first framework for connecting lifelong learning of students across multiple institutes and improve personalized learning by constructing and providing reusable/extendable user models. Our proposed framework focuses on constructing ontologies that represents a learner’s knowledge from their learning logs on learning platforms. This work solves the challenge with making sense of learning logs previously generated in a different learning platform. In figure 1, we show our proposed framework and how we solve this problem by: (A) designing processes for data collection, (B) model generation and unification and, (C) integration of a query-enabled interface to support learning. In the following subsections, we will discuss each of the components of the proposed framework.

4.1 Data Input and Collection

We first discuss the method for data input and collection from multiple sources. Because students may have engaged in learning on different platforms, it becomes necessary to define how their learning logs from these systems may be securely collected and held privately. Ocheja et al. (2018) already proposed
Figure 1. Connected lifelong learning framework.

Figure 2. Blockchain of learning logs (BOLL) data transfer processes.

4.2 The Connector

This stage involves processing learning logs from multiple learning systems and using the resulting data to construct a user model such as a knowledge graph. For example, a learner who attended School A may enroll in School B. The learner can ask for their records to be transferred from School A through the blockchain to School B. School B could then use these records to create a new model or insights to onboard the newly enrolled learner. The data from the different learning environment enables the realization of a single knowledge graph representing the learner’s knowledge (Lecailliez, Flanagan & Ogata, 2019). One major relevance of the resulting knowledge map is to help students reflect on past learning activities that are useful to current/future learning tasks. In figure 3, we show a hypothetical
knowledge map representing a learner’s lifelong learning from K-12 through to High Education over time that goes from the past on the left to the present on the right. At each level on the map and on each node (left to right), a learner can look back to reflect on their learning journey and possibly re-evaluate their preparedness for succeeding learning tasks (nodes) using various interactive objects on these nodes such as quizzes, and available associated learning contents.

Figure 3. A sample learner’s lifelong learning knowledge map.

The user model is expected to give an automatically exploitable representation of the learner’s knowledge. This involve the ability to answer some key questions about the learner’s learning history, assess a learner’s readiness for a learning activity, and provide a base template for personalization in a new or existing learning environment. Also, the ability to update the model as the user progresses is important as learning is a continuous process.

4.3 Learner Model Interface with Scenarios

This phase mainly provides an interface to interact with and use the user model to support a learner’s learning journey. It requires the development of a query-enabled interface for learners and teachers to access and assess a learner’s knowledge and consequently provide learning support. For example, a teacher may be faced with the task of diagnosing why a student has performed poorly in a given quiz. The lifelong learner model becomes useful in providing answers to these questions in a faster and more comprehensive manner.

The user model may also be useful for learning content designers who want to deliver a well-tailored and personalized learning contents. This is made possible by the model’s ability to provide a broad representation of the learner’s readiness for new learning contents or activities. This can be measured by evaluating the extent to which a learner has covered prerequisite topics.

Another scenario where a lifelong learner model may be useful is in helping students to adjust their learning habits by learning from other students. This particular feature can be enabled through the decentralized architecture of the BOLL system when learners on the blockchain can anonymously probe the system for useful information. For example, a student at a university may be interested in knowing what learning patterns are being adopted by high achievers in their class or at other schools.

5. Discussion

In the beginning, this work set out to solve some problems with realizing distributed lifelong learner models. Here, we will address each problem and evaluate how the proposed framework solves it.

The ubiquitous nature of learning makes it difficult but necessary to connect all learning activities in order to realize a lifelong learning. Our proposed framework solves the challenge with
assessing a learner’s knowledge by processing learning records of a learner from various systems and using such data to construct a model that represents the learner’s knowledge. Thus, the resulting model can be queried to understand what a learner knows. Because the constructed model relies on the learner’s lifelong learning, this model also becomes useful in assessing a learner’s lifelong learning.

The cold-start problem is usually faced by learning systems when such systems attempt to provide personalization to the user without prior information about the user. Our proposed framework solves the cold-start problem by enabling learning systems to integrate with the input arm (see A in figure 1) of the proposed framework so as to gain access to a learner’s previous learning records at other institutes. The learning system could then use such information to determine what kind of personalization to provide.

Effectively tracing learning difficulties provide potentials to solving such difficulties. The framework proposed in this paper enables diagnosis of learning difficulties by providing a mechanism to measure students’ knowledge on prerequisites topic. When a user’s knowledge in a prerequisite topic is below requirement, it becomes possible for their teacher to know what areas to places emphasis on during teaching/learning.

Our proposed framework enables students to share successful learning habits by allowing students to anonymous share their learning habits with other learners on the blockchain of learning logs. Students could in turn query the blockchain to gain access to useful information on their learning journey. The importance of using the blockchain to facilitate this is to ensure that students’ privacy rules are not violated and that only authorized entities can access protected information.

6. Limitations

We have shown how lifelong learning logs can be connected in a meaningful way across multiple systems with minimal information loss. Our proposal is made possible by the presence of a decentralized platform for connecting learning logs: the BOLL platform. However, some challenges exist.

The differences in learning logs format makes it difficult to process logs from different systems. While this work does not set out to solve problems with unifying logs defined in different standards, we acknowledge the existence of such a problem. It is therefore necessary for institutes to adopt standards that facilitate interoperability across different learning systems such as the xAPI standard.

Scalability of the blockchain is one of the main limitations to adopting the blockchain as extensively discussed by Vukolić (2015). The BOLL platform also suffers from the limitations of the proof of authority consensus algorithm upon which it runs. For instance, to add a learning log to the BOLL network will take from 15 seconds to 2 minutes (Ocheja et al., 2019). The problem with this is that learning logs are generated at a much faster rate than 15 seconds. Therefore, it is necessary to determine the best approach to write these logs on the BOLL network. One way we have identified to solve this problem, is to mine only representative learning logs to the blockchain and also to batch multiple learning logs in a single transaction. Initial experiments with this approach showed significant improvement. For example, over 1 million records which would take more than a year to write to the blockchain were transferred over a two weeks period using mining of representative learning logs.

7. Conclusion

In this work, we presented the current limitations with learning systems in connecting, and making sense of logs from multiple learning environments. To solve this problem, we proposed a framework that builds on the BOLL platform to construct models that can represent a learner’s knowledge. This work also presented some practical scenarios where such models can be used to support learning and we also discussed the implications.

8. Future Work

Future work will be focused on making a concrete implementation of the proposed framework and validating its effectiveness in a typical learning environment. We will continue with the participatory co-design process at every stage of the implementation to ensure that the outcome remains consistent
with the needs of the stakeholders. At the core of the proposed framework is the BOLL platform which we have been actively involved in its development. This makes it possible for us to be able to provide answers to the limitations of the BOLL platform identified in section 6. We also acknowledge that some aspects of the proposed framework may be revised during the course of the implementation to meet some system dynamics not initially anticipated. Consequently, we will provide an updated version (if any) of the framework in subsequent publications.

Acknowledgements

This work was partly supported by JSPS Grant-in-Aid for Scientific Research (B)20H01722, JSPS Grant-in-Aid for Scientific Research (S)16H06304 and NEDO Special Innovation Program on AI and Big Data 18102059-0.

References


Using Sequence Clustering to Unveil Students’ Learning Strategies and Explore the Relationship with Cognitive Load


*a Computer Science & Information Engineering, National Central University, Taoyuan City, Taiwan
b College of Computer Science, National Pingtung University, Pingtung City, Taiwan
c Information & Computer Engineering, Chung Yuan Christian University, Taoyuan City, Taiwan
d Department of Accounting, National Changhua University of Education, Changhua City, Taiwan
*anna.yuqing@gmail.com

Abstract: The generation of learning strategies is related to students' cognitive abilities, and students with similar cognitive abilities tend to adopt similar learning strategies. This study conducted an experiment of a system programming course with 113 college students and collected logs from a learning management system. There are five learning strategies that can be identified through sequence clustering analysis, they are comprehensive video-oriented learning, comprehensive slide-oriented learning, offline learning, selective learning, and reflective learning. Among the five learning strategies, selective learning strategy is positively correlated with students’ cognitive load, which can be explored that these students make practice of quiz without reading learning materials, and tend to select unfamiliar concepts to read based on quiz results. Students who use selective learning strategy are referred to as incomprehensive learners, and the selective learning strategy prone to increase students’ cognitive load due to overlying unclear and segmented information.

Keywords: Cognitive Load, Sequence Clustering, Learning Strategies

1. Introduction

There are some learning strategies or patterns will significantly affect learning performance, such as cognitive load (Paas, 1992), which makes it necessary to reveal the student ’s condition immediately. A large number of students’ interactions with computers can be summarized and extracted from the logs of the digital learning environment. Therefore, these logs that are generated in real time and accurately recorded may be used as an auxiliary mechanism for questionnaire measurement. Many studies have started from software logs, such as: Akçapınar et al. (2019) developed an early warning system for students by logs which generated from e-Book software. However, students' learning strategies is a potential structure and cannot be observed directly in learning logs, but instead must use the appropriate analysis methods and techniques of data mining, past research has shown that by converting students 'learning logs into sequences, and the learning strategies of students can be well found through sequence clustering (Akçapınar et al., 2020; Jovanović et al., 2017).

In summary, this study focuses on identifying a most suitable sequence clustering for extracting students’ learning strategy from learning logs which collected from learning management system (LMS). In addition, this study examined the correlation between students’ learning strategies and cognitive load. Therefore, there are two research questions in this study.

RQ1: How to extract students’ learning strategy from the logs collected from LMS?
RQ2: What is the correlation between students’ learning strategy and cognitive load?
2. Methodologies

2.1 Participants

This study is based on a system programming course at a university in Taiwan. The experiment duration ran for 9 weeks. There are 113 students participating in the course. Moreover, this course used a software named i-learning as the LMS platform, which includes teaching material videos, slides, discussion areas, teaching material downloads, and test.

2.2 Instrument

We adopted a cognitive load instrument for measuring students’ learning status. The number of valid questionnaires is 88 out of 113 students. It is used to evaluate the load generated by students’ cognition of the system program course. The cognitive load questionnaire in this study was based on Sweller et al. (1998) and Paas (1992). The questionnaire is a 6-point Likert scale, a total of eight questions, ranging from 1 (very disagree) to 6 (very agree). The cognitive load questionnaire contains two parts: questions 1-5 were a group as the first part named mental load, and questions 6-8 were a group as another part named mental effort.

2.3 Constructing students’ learning sequence

In order to perform sequence analysis, this session will introduce the process and method for students to extract the raw data from the i-learning LMS to the sequence, and collect students’ learning logs for 9 weeks. This study uses final exam score to represent students’ learning performance. Students’ learning logs on the i-learning LMS will be recorded in the database. Each event record is represented by three fields, including username, event begin time, and event title as shown in Table 1.

<table>
<thead>
<tr>
<th>Reclassified event category</th>
<th>Related event categories in this category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regulations</td>
<td>Information about course, grading standards, teaching objectives</td>
</tr>
<tr>
<td>Test</td>
<td>Events related to online test</td>
</tr>
<tr>
<td>Course key points video</td>
<td>Related key points videos by the teacher in course</td>
</tr>
<tr>
<td>Teaching material download</td>
<td>Course notes, reference materials, program execution examples, noun explanation</td>
</tr>
<tr>
<td>Previous exam</td>
<td>Previous exam by year</td>
</tr>
<tr>
<td>Course video</td>
<td>Video teaching materials for teachers' various chapters</td>
</tr>
<tr>
<td>Course slide</td>
<td>Slides of the teaching materials of the teachers' various chapters</td>
</tr>
<tr>
<td>Discuss</td>
<td>Course discussion, software development discussion, course content explanation</td>
</tr>
<tr>
<td>Questions related to the course</td>
<td>Learning review videos</td>
</tr>
</tbody>
</table>

2.4 Sequence clustering methods

We used agglomerative hierarchical clustering method to cluster students’ learning sequences. For each iteration of agglomerative hierarchical clustering method, the clustering task starts from the tree leaves and then combines two nearest clusters into one. This technique is considered particularly suitable for detecting student groups in online learning environments (Kovanović et al., 2015). Aggregated hierarchical clustering method starts aggregation from the bottom of the tree structure, and the algorithm steps are as follows (Murtagh, 1983):

1. At the beginning, each data point is a cluster denoted by $C_i$, $i = 1, 2, ..., n$.
2. Find the nearest two clusters $C_i$ and $C_j$ within all clusters.
3. Combine $C_i$ and $C_j$ into a new cluster.
4. If the number of clusters is equal to the desired one, stop. Otherwise go back to step 2 to create more clusters.

3. Results and discussion

The preprocessed sequence is clustered by agglomerative hierarchical clustering method, and the result shows the number of clusters is 5, and the smallest cluster size is 30 and the largest cluster size is 284. In order to answer the RQ1 (How to extract students’ learning strategy from the logs collected from LMS), we cluster a total of 621 learning sequences after pre-processing to detect students’ learning strategies on the LMS. The results of the sequence clustering show that students’ learning process will produce different types of learning strategies, which including:

**Comprehensive video-oriented learning strategy:** The learning sequence in this strategy is mainly composed of the interlaced reading of course video and course slide. Most of the first event in this learning sequence is watching the course video. We then sorted out the top three most frequently learning sequences in this learning strategy. The most common sequence was students stopped after course video, with a sequence length of 1, and the second most common sequence was to watch the course video followed by the course slide, the sequence length is 2, and we can find that the remaining sequences follow the course video connection to any event and then the course video, the sequence length is equal or larger than 3. From the above observation, we can consider that this learning strategies is relatively complete learning. Therefore, we named this strategy as comprehensive video-oriented learning.

**Comprehensive slide-oriented learning strategy:** The learning sequence in this strategy is mainly composed of course slide and course video. According to the statistics results, this strategy has higher number of the event of course slide than the strategy of comprehensive video-oriented learning. The average length of the learning sequence of this strategy is the longest than any other strategies. Most of the first event of this strategy are students will watch the course video. It can be seen from the above discussion that this learning strategies is similar as pervious one: starting with basic teaching material. However, we named this strategy as comprehensive slide-oriented learning due to the major chrematistic of this strategy is the slide-oriented learning.

**Offline learning strategy:** According to the statistics results of this strategy, it can be observed that the major sequences are teaching material download, which is considering as offline learning. Students will download teaching material at the first beginning of the class. Students in this strategy were search the chapter videos or slides that they want to watch. Therefore, we named this strategy as offline learning.

**Selective learning strategy:** According to the statistical results, it can be observed that the learning sequence in this strategy is mainly consists of previous exam. Most of the first event in this strategy is students tends to do previous exam. We further sorted out the top three most frequently learning sequences in this strategy and found that the most common sequence is to stop after doing the previous exam. The remaining two common learning sequences are to watch the course video after doing the previous exam, and to watch the course video after doing the previous exam. Both of remaining two common learning sequences' length are 2 or over 2. The selective learning strategy represents that students are reading for the purpose of being able to pass the course, and it can be found that the average number of times of the course video and the course slide in the sequence are lower than the overall average. This indicating that the selective learning strategy is lacks of learning from the fundamental concepts, considering that the behavior in this strategy is only for performance-oriented learning. Therefore, we named this strategy as selective learning.

**Reflective learning strategy:** The diversity of learning sequences in this strategy is relatively higher than other strategies. According to the statistical results, it can be observed that the proportions of the two events of course key points video and discussion are the highest in the learning strategy. This strategy means students will find key points to organize and participate in the discussion after watching the slide and video. It means that students have the ability to reflect on the teaching materials and can review the past experience to improve learning behavior. Therefore, we named this strategy as reflective learning.

To reply RQ2 (What is the correlation between students’ learning strategy and cognitive load?), we used Spearman correlation coefficients to explore the correlation between learning strategies and cognitive load. We test the significance of the correlation between the sum of the number of sequences
in each learning strategy of each student and cognitive load. The results of Spearman correlation coefficients indicated that the selective learning strategy has a significant positive correlation with mental load, while mental load refers to working memory load. The correlation is significant at the 0.05 levels, and the correlation coefficient value is 0.22. Cognitive load as the amount of load generated by applying specific tasks to students' cognitive systems. To further understand how selective learning strategy affects mental load, we test the significance of the correlation between the top three most frequently learning sequences of selective learning strategy and mental load.

Statistics results indicated that the mental load is significantly positively related to the sequence of watch the course video after doing the previous exam. The correlation is significant at the 0.05 levels, and the correlation coefficient value is 0.21. This sequence shows that the students tend to be performance-oriented first, and then find important information in the teaching material based on the previous exam, relying heavily on the exam content for learning, lack of basic concepts, students hope to receive important information about the exam in a short period of time. This learning strategy may cause students to be unable to load knowledge for a while, and according to past research, mental load is considered to be an important indicator that affects student' learning performance. We hope that by correcting this kind of learning strategy of the students to improve students' learning performance.

4. Conclusion and future research

After a series of log data processing processes, we have summarized five learning strategies for students: Comprehensive video-oriented learning, Comprehensive slide-oriented learning, Offline learning, Selective learning, and Reflective learning. The essence of the difference between the five strategies is the completeness of watching movies, learning materials, and the order of actions. The most critical one of the five strategies are Selective Learning. The main reason is that the strategy has a significant positive correlation with students' cognitive load. More specifically, when students skip watching videos and learning material and take quizzes directly, or the student takes the quiz first and then returns to the video or learning material to find the answers to the questions in the quiz; this is a symptom of the student's cognitive overload. The above summary clearly shows that in the design of learning activities or the design of teaching materials, to avoid the situation we mentioned above, it must induce or guide students to read at least once. Additionally, with a mechanism of encouragement to enable students to go back and reread the textbooks or videos under specific conditions, it will effectively reduce cognitive load and have the opportunity to improve learning outcomes. In the future, we will design and implement classroom activities under the concept mentioned above and encourage, guide and inducing students to read the content in a precise number of times through intervention or prevention mechanisms and expose the actual benefits to students during this process.

Acknowledgements

This work is supported by Ministry of Science and Technology, Taiwan under grants MOST-109-2511-H-008-007-MY3, MOST-108-2511-H-008-009-MY3, MOST-106-2511-S-008-004-MY3, and Ministry of Education, Taiwan.

References


Hiroyuki KUROMIYA*, Rwitajit MAJUMDARb, Taisyo KONDO*, Taro NAKANISHI*, Kensuke TAKII & Hiroaki OGATAb

aGraduate School of Informatics, Kyoto University, Japan
bAcademic Center for Computing and Media Studies, Kyoto University, Japan

*kiroyuki1993@gmail.com

Abstract: Recent spread of the COVID-19 forces governments around the world to have temporarily closed educational institutions. Although many studies were published to announce the best practice under the school closure, we need to understand the impact of school closure on students’ learning before that. In this paper, we evaluate the impact of the school closure on our online teaching-learning environment. We use CausalImpact model to infer the impact on our learning analytics system using the learning log stored in the system. The results show that the school closure increased the number of logs on LMS by 163%, but decreased the number of logs on e-book reader by 77%. However, focusing on a particular course, we found that students’ learning engagement on online system increased both in LMS and e-book reader. We discussed that it is caused by the following reasons: 1) Changes in major users on our online learning platform, and 2) Limited functions of our e-book reader which was developed for face-to-face learning, not online learning. Further, the results also suggested that CausalImpact model is useful for evaluating the effectiveness of a specific event from learning logs collected by learning analytics systems.

Keywords: Learning Analytics, School Closure, COVID-19, Time-Series Analysis, CausalImpact model, Secondary Education

1. Introduction

UNESCO reported that 90% of students are affected in some way by COVID-19 pandemic (UNESCO, 2020). Most governments around the world have temporarily closed educational institutions in an attempt to contain the spread of the virus. In Japan too, the national government requested temporary school closure for elementary, junior-high, and high school on March 2. For the educational stakeholders, there is an urgent need to respond to this situation as soon as possible. To respond to the needs, many studies were published to announce the best practice under the school closure (Dai & Lin, 2020; Morgan, 2020). However, at the same time, we need to understand the impact of school close on students’ learning. Without knowing the current situation, we cannot take measures. To respond these requests, this paper aims to investigate the impact of school close on the online teaching and learning environment at a secondary school in Japan. Our contributions of the paper are follows:

1. As we (our university) have been cooperating with a secondary school even before the disaster, we can directly assess the impact of school closure.

2. A specific time-series model, CausalImpact model was implemented for the first time to evaluate the impact in the educational sector.

The context of our study was a public secondary school in the Kyoto city of Japan. The school had 360 students at junior-high level and 845 students at the high-school level. The junior high students were provided a tablet from the school and high school students have their own tablet or PC. Although the national government requested to close the school on March 2 (Nikkei Asian Review, 2020), the school was actually closed from March 2. During online sessions during emergency period, students access their home internet connection.
In the next section, we will review our online learning platform and time series models used for the analysis.

2. **Foundation of current study**

2.1 Learning Evidence Analytics Framework (LEAF) and learning logs

We offer an integrated online learning platform called LEAF - Learning Evidence Analytics Framework to a secondary school in Japan (see Figure 1). LEAF consists of three learning support systems - Moodle, BookRoll, and LA view (Flanagan & Ogata, 2018). The first component, is any learning management system (LMS). Moodle is used to host the courses, manage learning resources, and conduct assessments to evaluate students’ learning. The second component, BookRoll is an e-book reader, where teachers upload learning materials for students. BookRoll has many advantages compared to just uploading reading materials as a pdf format to Moodle. Teachers can create quizzes and recommendations on the material. Students can write handwritten memos, highlight text, and answer questions on any device which has a browser. Moreover, the interaction logs are stored in the Learning Record Store (LRS) so that teachers can review students’ learning activities. LAView is the learning analytics dashboard that visualizes the learner interactions. Teachers can see students’ markers, memos, bookmarks, time spent on each page, and other reading behaviors.

![Figure 1. Learning Evidence Analytics Framework (LEAF).](image)

2.2 *Time Series Analysis*

To evaluate the impact of school closure on our learning platform, we take a time-series analysis approach to the log data stored on the system. We compared two different time periods - before and after the school closure. However, comparing two periods is difficult in time-series data. We cannot apply group-comparison methods such as t-test because time-series data are not independent samples over time. To solve this problem, we examined three time-series analysis models - Interrupted Time-Series, Prophet, and CausalImpact - which are able to deal with the comparison of two time periods.

The first candidate was Interrupted Time Series (ITS) model. ITS model is a study design which evaluates the effectiveness of population-level interventions (Bernal et al., 2017). Mathematically it is the segmented regression model with dummy variables representing the period of the intervention. The advantage of ITS model is its simplicity. As its structure is similar to linear regression models, we can easily interpret the model. However, its prediction power is weak due to the simplicity. Although our purpose is not prediction, it may cause problems in causal inference because of insufficient model flexibilities.

The second candidate was Prophet model developed by Facebook research (Taylor & Letham, 2018). Prophet model is a kind of Generalized Additive Model, which consists of three main
components - trend, seasonality, and holidays. The advantages of prophet model are that we do not have to extrapolate missing values because it takes time-series analysis as a curve fitting problem and it is proved to have high-accuracy to predict future values. However, our purpose of using statistical model is not to predict future, but evaluate the impact of a specific event. Sometimes it is difficult to infer the effectiveness of an event by Prophet model.

The last candidate was CausalImpact model proposed by (Brodersen et al., 2015). Unlike Prophet model, it is developed to evaluate the impact of a market intervention. It applies Difference-In-Difference (DID) concept to infer the causality from observational data. It offers theoretical evidence of evaluating causal impacts to the model. CausalImpact model evaluates the impact of an intervention \( I \) by following equations:

\[
I = \frac{1}{t-n} \sum_{t'=n+1}^{t} \phi_{t'}(\tau) \quad (1)
\]

where \( \phi_{t'}(\tau) = y_{t'} - \hat{y}_{t'}(\tau) \quad (2) \)

In the equation (1), we assume that the current time is \( t \) and the intervention happened at time \( n \). In equation (2), \( y \) is the dependent variable and \( \hat{y} \) is the estimated value by the model. Hence, CausalImpact model compute the impact of the intervention by taking the difference between the predicted value \( \hat{y}_{t'}(\tau) \) and the actual value \( y_{t'} \) to make counterfactual inference.

We summarized three models in Table 1. Based on the comparison of the three models, CausalImpact model was selected to be implemented in this study for the following reasons:

1. Causal Impact model specializes in evaluating the impact of an event while other models does not focus on causal inference problem.
2. Both ITS model and CausalImpact model are used for decision making rather than predicting future values, but CausalImpact model can handle with more types of data.

Next section, we will explain the data and our analysis flow of it.

### Table 1

<table>
<thead>
<tr>
<th>Type of Model</th>
<th>Interrupted Time Series</th>
<th>Prophet</th>
<th>CausalImpact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose</td>
<td>Prediction, Decision Making</td>
<td>Prediction</td>
<td>Decision Making</td>
</tr>
<tr>
<td>Characteristics</td>
<td>Simplicity</td>
<td>Easy to adjust with domain knowledge</td>
<td>Difference-In-Difference approach</td>
</tr>
</tbody>
</table>

3. **Methodology**

We conducted two analyses for evaluating the impact of the school closure (see Figure 2). All the log data are collected from Learning Record Store on the LEAF platform.

First, we evaluate the impact on the whole system, looking at the change of accumulated number of logs for each learning tool - Moodle, BookRoll. In this analysis, the logs between April 1, 2019 and April 15, 2020 were considered. We took the number of logs per day as the indicator.

Second, we focused on a specific course log in order to explore the change of their teaching-learning style further. It was an English course in high school for grade one students. The number of teachers and students enrolled in this course were 19 and 280 for each. The reason why we selected the course was that the course was one of the most active courses in Moodle during the school close period and they used both Moodle and BookRoll from the beginning. In this analysis, the logs between April 1, 2019 and March 31, 2020 were considered. We conducted a brief interview to a teacher in the course after analyzing the logs. We took average time spent per student as the indicator.
4. Results

4.1 The Impact on the Whole System

We evaluated the impact of school closure to the whole system in terms of number of logs per day. Figure 3 shows the descriptive plot of number of logs in Moodle and BookRoll. Table 2 shows the results of the CausalImpact model fitting. The results showed that the school closure increased the number of logs on Moodle by 163% while decreased the number of logs on BookRoll by 77%. Both confidence intervals (95% CI) were also significant.

Table 2

<table>
<thead>
<tr>
<th></th>
<th>Moodle</th>
<th>BookRoll</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute Effect</td>
<td>+2525 logs</td>
<td>-10,069 logs</td>
</tr>
<tr>
<td>95% CI</td>
<td>[1911, 3180]</td>
<td>[-17823, -2678]</td>
</tr>
<tr>
<td>Relative Effect</td>
<td>+163%</td>
<td>-77%</td>
</tr>
<tr>
<td>95% CI</td>
<td>[123, 205]</td>
<td>[-124, -19]</td>
</tr>
</tbody>
</table>

Figure 3. Number of logs on Moodle (left) and BookRoll (right) before and after school close. The gray area represents the school close period.

4.2 Impact on the English Course
Next, we evaluated the impact of school closure to the specific course in terms of students’ time spent on a day. Figure 4 shows the descriptive plot the amount of time spent in Moodle and BookRoll over the period. We can see that students’ learning activities on our online learning platform was more active in the school close period than the normal period. Table 3 shows the results from CausalImpact model. It shows that students’ time spent increased in both Moodle and BookRoll (2227% and 875% for each). The confidence intervals (95% CI) are also significant. According to the teachers, the class using our platform increased from once a week to four times a week during the school close period. These results matched to the situation.

Table 3

<table>
<thead>
<tr>
<th>Absolute Effect (min)</th>
<th>Moodle</th>
<th>BookRoll</th>
</tr>
</thead>
<tbody>
<tr>
<td>95 % CI</td>
<td>[2.5, 3.0]</td>
<td>[1.3, 1.9]</td>
</tr>
</tbody>
</table>

The teacher also noted that they started use more function in Moodle such as quizzes and feedback features and upload their original reading materials on BookRoll. Before the school closure, they had used Moodle just as a place to distribute learning materials and they only uploaded the commercially available teaching materials on BookRoll.

5. Discussion

5.1 Cause of the Results

So far, we found that in the school close period, 1) the number of Moodle log increased, while the number of BookRoll log decreased on the whole online learning system, but 2) focusing on a single course, there is a course where time spent on both Moodle and BookRoll had been increased. We expect that the results were caused by the following reasons:

1. Before the school closure, our online learning system was mainly used by junior-high teachers, but after the school closure, the main users were high school teachers. While junior-high teachers were well trained to use BookRoll because they had many training sessions, high school teachers were not.
2. The teachers may think BookRoll was difficult to use BookRoll in online teaching activities. Before the school closure, BookRoll was used in face-to-face learning context. They tend to use videos during the school close period. A previous study (Shatakshi & Nardev, 2020) also pointed out that the most common source of the online classes was “PPTs with Audio”, followed by “Videos.”

5.2 Novel Application of Analytical Process

Here, we would like to discuss about the applicability of our analysis process as well. CausalImpact model has been used mainly in engineering field e.g. the impact of the cyber policy activation on cyber-attacks (Kumar, Benigni, & Carley, 2016) or the implementation of product modularity on Bus manufacturing (Piran, Lacerda, & Camargo, et al., 2017). To the best of our knowledge, this is the first case of applying CausalImpact model in learning analytics field. Time-series analysis with causal inference theory will be important in learning analytics field. We plan to implement automated intervention evaluation module from learning logs in LEAF platform. It will be useful for teachers who want to review their teaching.

6. Conclusion

In this paper, we presented a process of evaluating the impact of the school closure on the online learning system. We used the CausalImpact model to analyze students’ learning logs. The results suggested that the teaching-learning style had changed before and after the school closure drastically and the CausalImpact model is useful for evaluating the effectiveness of some events from learning logs collected by learning analytics systems. Moreover, time-series analysis with causal inference theory will be important in the next learning analytics researches.

Acknowledgements

This research was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (S) Grant Number 16H06304 and JSPS KAKENHI Grant-in-Aid for Research Activity Start-up Grant Number 18H05746.

References

Improving Classification in Imbalanced Educational Datasets using Over-sampling

Zeynab MOHSENI*, Rafael M. MARTINS*, Marcelo MILRAD* & Italo MASIELLO*
*Department of Computer Science and Media Technology, Linnaeus University, Sweden
*zeynab.mohseni@lnu.se

Abstract: Learning Analytics (LA) involves a growing range of methods for understanding and optimizing learning and the environments in which it occurs. Different Machine Learning (ML) algorithms or learning classifiers can be used to implement LA, with the goal of predicting learning outcomes and classifying the data into predetermined categories. Many educational datasets are imbalanced, where the number of samples in one category is significantly larger than in other categories. Ordinarily, it is ML’s performance on the minority categories that is the most important. Since most ML classification algorithms ignore the minority categories, and in turn have poor performance, so learning from imbalanced datasets is really challenging. In order to address this challenge and also to improve the performance of different classifiers, Synthetic Minority Over-sampling Technique (SMOTE) is used to oversample the minority categories. In this paper, the accuracy of seven well-known classifiers considering 5 and 10-fold cross-validation and the F1-score are compared. The imbalanced dataset collected based on self-regulated learning activities contains the learning behaviour of 6,423 medical students who used a web-based study platform—Hypocampus—with different educational topics for one year. Also, two diagnostic tools including Area Under the Receiver Operating Characteristics (AUC-ROC) curves and Precision-Recall (PR) curves are applied to predict probabilities of an observation belonging to each category in a classification problem. Using these diagnostic tools may help LA researchers on how to deal with imbalanced educational datasets. The outcomes of our experimental results show that Neural Network with 92.77% in 5-fold cross-validation, 93.20% in 10-fold cross-validation and 0.95 in F1-score has the highest accuracy and performance compared to other classifiers when we applied the SMOTE technique. Also, the probability of detection in different classifiers using SMOTE has shown a significant improvement.

Keywords: Learning Analytics, Imbalanced Dataset, Machine Learning, SMOTE, ROC, PR.

1. Introduction

The data generated in educational environments such as courses that use learning management systems or digital learning materials are frequently large, complex and heterogeneous (Martins et al., 2019). Learning Analytics (LA) (Khosravi & Cooper, 2017) can be defined as the measurement, collection, analysis and reporting of data about learners and their contexts and contains a wide range of methods for understanding and optimizing learning and the digital environments in which it occurs. LA is commonly implemented with the use of different ML algorithms (Chung & Lee, 2019), which are strong and flexible methods to produce solutions for problems that are not being handled with traditional statistical approaches. ML algorithms make it possible to predict students’ performance and risk of under- or over-performing, for example, as it helps to classify the data into predetermined categories (e.g., high performance, medium performance, low performance) (Akcapinar et al., 2019; Khosravi & Cooper, 2017).

Large-scale datasets usually contain several categories, but when the distribution of such categories is not uniform, i.e. some of them (the minority) are heavily under-represented when compared to others (the majority), that leads to category imbalance. This ‘imbalanced’ or ‘skewed’ distribution of category instances results in learning classifiers being biased (Kuncheva et al., 2019; Napierala & Stefanowski, 2016). One possible solution to this problem is over-sampling, where randomly-selected samples from the minority categories are duplicated. SMOTE is one of the most
popular algorithms for over-sampling, relying on the concept of nearest neighbours to create its synthetic data (Fernandez et al., 2018).

In this paper, we describe the improvement of the performance results of different ML classification algorithms using the SMOTE resampling techniques, applied to a learning analytics problem and dataset. Our research question is formulated as follows: How does the combination of SMOTE and under-sampling perform, in comparison to traditional ML classification algorithms, when handling learning analytics datasets? In our data, the activity of each student reflects the number of questions she/he answered related to a certain subject or topic. We analyse and compare seven different types of ML algorithms using two threshold metrics, accuracy and F1-score, together with two rank metrics, AUC-ROC curves and PR curves (Jeni, Chon & De La Torre, 2013).

The rest of this paper is organized as follows. Section 2 describes briefly related work in this field. Sections 3 and 4, respectively, present the case study and the different methods used in our work. An exploratory analysis and the experimental results are described in sections 5 and 6, respectively. Section 7 discusses the conclusions and presents possible lines of future work.

2. Related Work

Imbalanced classification is a problem in data sets with skewed distributions of data points (Chawla et al., 2002; Napierala & Stefanowski, 2016). Data-level approaches are used to address the category imbalance problems. The focus of these methods is on re-sizing the training datasets to balance different labels and make the dataset suitable for a standard learning algorithm. In order to fix category imbalance, resampling methods such as under-sampling and over-sampling methods are used. In under-sampling techniques, samples from the majority category are discarded, while in over-sampling methods, new minority category samples are generated (Mathew et al., 2015). SMOTE, an approach proposed by Chawla et al., (2002), is one of the most well-known over-sampling algorithms. It generates new minority data instances by identifying nearest neighbours in input space and applying a linear interpolation between them. This way, the new data instances populate areas near other points, and should properly resemble real data. Batista et al., (2004) have performed a systematic experimental study with 15 real-world datasets and different pre-processing methods such as SMOTE. The results indicate that the over-sampling methods provided better AUC than the under-sampling ones. Mathew et al., (2015), on the other hand, proposed a kernel-based SMOTE algorithm that generates synthetic data points of minority categories directly in the feature space of Support Vector Machine (SVM) classifier. In Beyan & Fisher (2015), a new hierarchical decomposition method for imbalanced data sets is proposed. The proposed method is based on clustering and outlier detection. These works are focused on general classification problems for imbalanced datasets with the goal of improving the performance of different algorithms but are not applied to the specific challenges and opportunities of imbalanced learning analytics applications. Hasnine et al., (2018) use SMOTE as one of the pre-processing steps in their pipeline, among others such as features selection, in order to apply ML for the prediction of student performance. However, their paper does not focus on the analysis of SMOTE itself, so it is not clear what exactly the advantages and disadvantages of the technique are when applied to an imbalanced learning analytics dataset. In our paper, we perform an experimental comparison that is specifically suited to compare performances before and after SMOTE, isolating its effects and producing objective insights on the impact of its application. By analysing the PR and ROC-AUC curves obtained from different classification algorithms, more in-depth understanding of the behaviour of balancing methods is provided.

3. Dataset Description

We used an imbalanced dataset which includes the learning behaviours of 6,423 medical students (data points) who used an online study platform (Hypocampus\textsuperscript{1}) during one year. The students chose their own study path through the material, which is arranged in subjects, e.g., orthopaedics or neurology, and

\textsuperscript{1} https://www.hypocampus.se/.
topics (or chapters), e.g., cerebrovascular disease or diabetes. During their studies, they are frequently faced with questions to test what they have learned. We aggregated all students’ answers per topic in order to generate the features, resulting (after data cleaning and pre-processing) in 1,445 features which reflect how many questions she/he answered on each of the 1,445 topics (Martins et al., 2019). In summary, each of the 6,423 data points indicate the learning behaviour of one student, quantified by the amount of questions she/he answered in each of 1,445 available topics. Additionally, each student may have a University ID, which indicates where that student comes from (or "Other" for all the students who do not have a university ID).

4. Methods

In order to identify how well a classifier preformed, a cross-validation procedure was used. In k-fold cross-validation, a partition of the dataset is formed by splitting it into k non-overlapping subsets, including k-1 training sets and one test set. Then, we can train and test the model k times, each time using different train and test sets (Goodfellow et al., 2016; Geron, 2017). We used 5-fold and 10-fold cross-validations; 5-fold means for each classifier we choose the mentioned input features that perform best on average when we train on 80% of the data and test on the remaining 20%, and 10-fold when we train on 90% of the data and test on the remaining 10%. To compare cross-validation results from different classifiers, one of the measures used is the average accuracy (or average error), shown in Eq. 1 (Chawla et al., 2002). In this equation TP or True Positives is the number of positive examples correctly classified; TN or True Negatives is the number of negative examples correctly classified; FP or False Positives is the number of negative examples incorrectly classified as positive; and FN or False Negatives is the number of positive examples incorrectly classified as negative.

Error rate, that is, \( I-Accuracy \), is more appropriate to use in balanced datasets, while other measures such as ROC and PR curves are more suitable to be used when there are unequal error costs. Using PR curves is more suitable for highly-skewed domains where ROC curves may provide an excessively optimistic view of the performance (Chawla et al., 2002). In this paper, two ROC-AUC and PR curves were used to compare different classification algorithms, summarized with the average precision, micro-average and macro-average. A macro-average computes the metric independently for each category, and then takes the average, whereas a micro-average aggregates the contributions of all categories to compute the average metric. In ROC-AUC and ROC curves, the True Positive Rate (Eq. 2) is a fraction calculated as the total number of true positive predictions divided by the sum of the true positives and the false negatives, while the False Positive Rate (Eq. 3) is calculated as the total number of false positive predictions divided by the sum of the false positives and true negatives. In PR curves, Recall (Eq. 4) is a metric that quantifies the number of correct positive predictions made out of all positive predictions that could have been made. Recall is calculated as the number of true positives divided by the total number of true positives and false negatives. In this curve, Precision (Eq. 5) is a metric that quantifies the number of correct positive predictions made, and it is calculated as the number of true positives divided by the total number of true positives and false positives (Chawla et al., 2002). Finally, to determine a weighted average of the precision and recall values, \( F_{1}\text{-score} \) is used (Eq. 6) (Jeni, Chon & De La Torre, 2013). \( F_{1}\text{-score} \) range is between 0 and 1, where the maximum shows the perfect classification.

\[
\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad \text{(Eq.1)}
\]

\[
\text{TP} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \text{(Eq.2)}
\]

\[
\text{FP} = \frac{\text{FP}}{\text{TN} + \text{FP}} \quad \text{(Eq.3)}
\]

\[
\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \text{(Eq.4)}
\]

\[
\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad \text{(Eq.5)}
\]

\[
\text{F}_{1} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}} = \frac{2 \times ((\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}))}{\text{(Eq.6)}}
\]

Additionally, in order to compare and check the statistical significance of the difference in results before and after the use of SMOTE, we ran a set of one-tailed Mann-Whitney U tests, one for each performance measure, with the significance level set to 0.01. These are non-parametric statistical tests suitable for small samples, as is our case.
4.1 Classification Algorithms

We compare seven different types of ML classification algorithms: Linear (Geron, 2017), k-Nearest Neighbors (kNN) (Goodfellow et al., 2016), Decision Tree (Goodfellow et al., 2016), Neural Network (Geron, 2017), Support Vector Machines (SVM) (Geron, 2017), Random Forest (Breiman, 2001) and XGBoost (Geron, 2017). Due to space limitations, and the fact that these are well-known in the ML field, we refrain from discussing them in details here; please see the references for more information. Each of these seven well-known algorithms can be configured with different hyperparameters that will affect their performance; therefore, we report only on the best configurations found after performing a systematic hyperparameter search in each case.

5. Experimental Results

Table 1 contains the results of all the considered metrics for the seven ML algorithms, before applying the SMOTE method. The mean performance of all algorithms stays around 58%, and while micro-average ROC is relatively high (0.85), the rest of the measures including average precision (0.56), macro-average ROC (0.71) and F1 scores (0.57) reflect the expected low-quality results of an imbalanced dataset. The Random Forest classifier had the best results overall, with 61.70% accuracy in 5-fold cross-validation, 61.79% in 10-fold cross-validation, 0.63 average area in PR, 0.88 micro-average area, 0.76 macro-average area in ROC-AUC, and F1 score of 0.62. It was the strongest classifier before the application of SMOTE.

Table 2 shows the results for the same algorithms after applying SMOTE to correct the imbalance of the data. The overall results improved drastically, with mean accuracy now around 78% in all cases, and all the rest of the measures above 0.80 (mean). All algorithms showed increased performance overall, with the Neural Network and Random Forest classifiers staying at the top with 92.77% and 92.40% accuracy in 5-fold cross-validation, and with 93.20% and 92.81% in 10-fold cross-validation, respectively. Random Forest also remains the best classifier according to the rest of the performance measures, with 0.96 average area in PR, 0.98 micro-average area, and 0.98 macro-average area in ROC-AUC. Neural Network, with 0.95, has the highest F1 score, followed closely by Random Forest with 0.94. It is interesting to notice that Linear and XGBoost did not improve much after SMOTE, reflecting the highly non-linear and complex nature of the data (to which these algorithms are not suitable).

The last row of Table 2 provides the U-values resulting from the Mann-Whitney U tests for each measure (as described in Section 4). In all the considered cases, the critical value of $U$ at $p < .01$ is 6, so a U-value of less than 6 means statistically-significant results. Most of the measures show a statistically-significant increase in the overall performance (highlighted in green), except Micro-average ROC (highlighted in red).

To further illustrate and discuss these results, we focus on the Random Forest algorithm (which performed best overall) and show, in Figure 1, three detailed visual comparisons of its performance: the confusion matrix, as a heatmap (Pryke, Mostaghim & Nazemi, 2007); the ROC curves; and the PR curves. In the confusion matrix (Figures 1a, d), darker cells mean correct class predictions. The main problem with imbalanced datasets is immediately apparent in the matrix before SMOTE (Figure 1a): the classifier assigned the label “other” to most points (since most darker cells are in the last column to the right), resulting in low performance. After SMOTE, however, the problem is mostly solved, as can be seen from the darker cells along the diagonal of the matrix. In the ROC curves, the ideal results are curves that bend along the top-left of the graph, maintaining a large proportion of TP vs. FP, as is the case after SMOTE (Figure 1e). Figure 1b shows that, before SMOTE, the lines are close to the diagonal instead. On the other hand, for the PR curve graph, the ideal results are curves that are close to the top-right corner, as is again the case after SMOTE (Figure 1f). In Figure 1c we can see that, before SMOTE, the lines were near random and scattered among the whole graph.
Table 1. Performance of ML Algorithms Before SMOTE

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>Training (%)</th>
<th>5-fold (%)</th>
<th>10-fold (%)</th>
<th>Avg. Precision</th>
<th>Micro-avg. ROC</th>
<th>Macro-avg. ROC</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>54.47</td>
<td>54.29</td>
<td>54.99</td>
<td>0.47</td>
<td>0.80</td>
<td>0.69</td>
<td>0.48</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>57.74</td>
<td>57.64</td>
<td>58.23</td>
<td>0.55</td>
<td>0.85</td>
<td>0.70</td>
<td>0.56</td>
</tr>
<tr>
<td>Neural Network</td>
<td>53.07</td>
<td>54.85</td>
<td>54.65</td>
<td>0.52</td>
<td>0.84</td>
<td>0.73</td>
<td>0.52</td>
</tr>
<tr>
<td>SVM</td>
<td>59.77</td>
<td>59.74</td>
<td>59.91</td>
<td>0.58</td>
<td>0.85</td>
<td>0.69</td>
<td>0.59</td>
</tr>
<tr>
<td>kNN</td>
<td>60.16</td>
<td>60.02</td>
<td>59.94</td>
<td>0.59</td>
<td>0.85</td>
<td>0.72</td>
<td>0.59</td>
</tr>
<tr>
<td>Random Forest</td>
<td>61.95</td>
<td>61.70</td>
<td>61.79</td>
<td>0.63</td>
<td>0.88</td>
<td>0.76</td>
<td>0.62</td>
</tr>
<tr>
<td>XGBoost</td>
<td>62.02</td>
<td>61.30</td>
<td>61.31</td>
<td>0.61</td>
<td>0.86</td>
<td>0.74</td>
<td>0.61</td>
</tr>
<tr>
<td>Mean</td>
<td>58.45</td>
<td>58.50</td>
<td>58.69</td>
<td>0.56</td>
<td>0.85</td>
<td>0.71</td>
<td>0.57</td>
</tr>
</tbody>
</table>

Table 2. Performance of ML Algorithms After SMOTE

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>Training (%)</th>
<th>5-fold (%)</th>
<th>10-fold (%)</th>
<th>Avg. Precision</th>
<th>Micro-avg. ROC</th>
<th>Macro-avg. ROC</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>64.72</td>
<td>64.15</td>
<td>64.56</td>
<td>0.58</td>
<td>0.82</td>
<td>0.82</td>
<td>0.59</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>84.09</td>
<td>84.24</td>
<td>85.11</td>
<td>0.71</td>
<td>0.75</td>
<td>0.74</td>
<td>0.71</td>
</tr>
<tr>
<td>Neural Network</td>
<td>92.75</td>
<td>92.77</td>
<td>93.20</td>
<td>0.93</td>
<td>0.88</td>
<td>0.88</td>
<td>0.95</td>
</tr>
<tr>
<td>SVM</td>
<td>60.31</td>
<td>61.04</td>
<td>61.64</td>
<td>0.85</td>
<td>0.92</td>
<td>0.92</td>
<td>0.86</td>
</tr>
<tr>
<td>kNN</td>
<td>83.69</td>
<td>84.24</td>
<td>84.85</td>
<td>0.89</td>
<td>0.97</td>
<td>0.97</td>
<td>0.9</td>
</tr>
<tr>
<td>Random Forest</td>
<td>92.07</td>
<td>92.40</td>
<td>92.81</td>
<td>0.96</td>
<td>0.98</td>
<td>0.98</td>
<td>0.94</td>
</tr>
<tr>
<td>XGBoost</td>
<td>69.61</td>
<td>68.29</td>
<td>68.60</td>
<td>0.75</td>
<td>0.89</td>
<td>0.89</td>
<td>0.75</td>
</tr>
<tr>
<td>Mean</td>
<td>78.17</td>
<td>78.16</td>
<td>78.68</td>
<td>0.81</td>
<td>0.89</td>
<td>0.88</td>
<td>0.81</td>
</tr>
</tbody>
</table>

U-values 2 (<6) 2 (<6) 1 (<6) 3.5 (<6) 13.5 (>6) 1.5 (<6) 3 (<6)

Figure 1. Confusion matrices, ROC curves, and PR curves of Random Forest classifier before (a, b, c) and after (d, e, f) the use of SMOTE.
6. Conclusion

In this article we have compared the performance of seven machine learning algorithms applied to an imbalanced LA problem, to answer our research question on how the combination of SMOTE and under-sampling performs compared to traditional ML classification algorithms. The students' dataset was collected from a Web-Based Learning Environment during one year and it consists of students (data points) described by multidimensional numerical vectors (features). The approach described here should be generalizable to any other scenario similar to this. According to the results, the performance using SMOTE has widely increased, and Neural Network and Random Forest are the most accurate and high-performance classifiers among the tested ML classification algorithms. Thus, we determine that the combination of SMOTE and under-sampling performs better than traditional ML classification algorithms in an LA context, which reflects other previous and more general results outside of LA. One limitation of this approach is that SMOTE is based on linear interpolation between nearest neighbours, which limits its application for datasets that are too large and contain highly non-linear relationships between its features. The results of a high-performance classification algorithm on educational datasets can have practical implications for teachers, that is, given the right visualization technique this sort of analysis promise to guide teachers in identifying learning issues and possibly, in the future, predicting students’ outcomes.

References


Applying Key Concepts Extraction for Evaluating the Quality of Students’ Highlights on e-Book

Albert YANG*, Irene Y.L. CHENb, Brendan FLANAGANC & Hiroaki OGATAc

*Graduate School of Informatics, Kyoto University, Japan  
bDepartment of Accounting, National Changhua University of Education, Taiwan  
cAcademic Center for Computing and Media Studies, Kyoto University, Japan  
*yang.ming.35e@st.kyoto-u.ac.jp

Abstract: The quality of students’ highlights can be an indicator of their learning performance. While the most common approach to grade their highlights is by humans, human grading can be inconsistent, especially when the number of highlights are large or when graders have different background knowledge. In this research, we propose a model to automatically extract important concepts from class materials, analyze students’ highlights and find the correlation between highlight quality and students’ learning performance. We first compared different text summarization algorithms with different evaluations to see which of them generates the summarization that is closest to the reference answer generated by humans. Then we used the selected algorithm to summarize the text from learning materials as important concepts, and compared the summaries with students’ highlights to calculate their highlight scores. Finally, we considered the highlight score from the best method as the highlight quality and observed whether it has a correlation to students’ learning performance.

Keywords: E-learning, text summarization, learning analytics

1. Introduction

In this study, we analyzed students’ highlight records of a 12-week course in a university. There were 44 students enrolled in this course. A total of 22 slides of learning materials were uploaded to BookRoll. BookRoll is a digital learning material (e-Book) reading system (Flanagan & Ogata, 2017) that provides students with e-Books and records their behaviors while reading. Students’ e-Book reading actions in BookRoll have been described in detail by (Ogata et al., 2015). Before each class, the instructor uploaded the slides of the class to BookRoll for students to preview. Students were asked to highlight the words or sentences they think are important using the marker function on BookRoll. There are two types of markers they can use, red marker for important concepts and yellow marker for concepts they feel difficult to understand. The instructor provided a reference answer which is the key concepts highlighted by the instructor every week. We then used this reference answer to test different summarization techniques, evaluate students’ highlights, and find the relationship between highlight quality and students’ learning performance. This study aims to answer the following research questions:

RQ1: What is the best key concepts extraction algorithm for evaluating the quality of students’ highlights on e-Book?

RQ2: Does the quality of students’ highlights affect their learning performance?

2. Method

2.1 Preprocessing

Since the content of the slides uploaded by the instructor is in PDF format, it needs to be converted to
raw text for our analysis. We used python's pdfminer package to convert the content to a plain text file. Both instructor and students' highlights were collected in text form using BookRoll. We applied preprocessing techniques on the raw text of slides and highlights, such as removing special characters and converting text to lowercase. In addition, since the students may delete a marker after adding it, there are two types of records in the e-Book, ADD_MARKER and DELETE_MARKER. In order to ensure that the records we use are the highlights reserved by the students, we ignored the highlights which exist in both records.

2.2 Methods for Key Concepts Extraction

After preprocessing the content of the slides and the highlights, we extracted the important concepts from the slides using text summarization algorithms. TextRank (Mihalcea & Tarau, 2004) and RAKE (Rose et al., 2010), which belong to traditional machine learning methods, and BERT (Devlin et al., 2018), a deep learning architecture were used in this study.

For TextRank, we cut the text into multiple sentences, and selected a quarter of sentences as a threshold to represent the key concepts. TextRank can be used to extract either keywords or key sentences from text.

For RAKE, the full text was passed to the model and a quarter of phrases were extracted as the threshold to represent the key concepts.

For BERT, two approaches were used to tokenize the class materials. We tokenized the text into sentences and pages. To cut the text into pages, we concatenated every sentence in each page as a long sentence. Our approach is to tokenize the incoming text into sentences or pages, pass the tokenized sentences or pages to BERT for inference to output embeddings, and then cluster the embeddings with K-Means. Since the embeddings include more than 700 dimensions which cannot be passed directly to K-Means, we applied PCA to reduce the number of dimensions to two. Centroids of the clusters in the vector space represent the key concepts in the original text. For each key concept, we selected the embedding sentence or page that is closest to the centroid as the represented sentence or page. Figure 1 shows the result of concepts clustering for slides using K-Means.

![Figure 1](image)

*Figure 1.* Using K-Means to cluster the embedding sentences (left) and pages (right). Dark centroids represent the key concepts in the slides. Each points represents a sentence (left) or a page (right). X axis and Y axis are two principle components generated by PCA.

After extracting key concepts from the learning materials, we used the highlights provided by instructor as reference answer to evaluate the quality of summaries from machine using BLEU 1, BLEU 2, BLEU 3, BLEU 4 (Papineni et al., 2002), METEOR (Denkowski and Lavie, 2014), and ROUGE (Lin, 2004).

3. Results and Discussion

For Research question 1, What is the best key concepts extraction algorithm for evaluating the quality of students’ highlights on e-Book?

We tested TextRank, RAKE, and two variants of BERT as different text summarization
techniques and compared the summaries with reference answers using standard metric BLEU, METEOR and ROUGE. The best results were obtained using the traditional machine learning method, RAKE. It outperforms other three methods in all metrics except BLEU 4 and ROUGE-L. This is expected as the summaries of RAKE are phrases while others are sentences. BLEU 4 takes into account the 4-gram co-occurrence while the length of phrases generated by RAKE is often shorter than the sentences generated by other methods. Similarly, ROUGE-L measures the longest common sequence which is also bad for RAKE. The results obtained for the different algorithms we used are showed in Table 1.

Table 1. Evaluation results for the four text summarization techniques.

<table>
<thead>
<tr>
<th></th>
<th>BLEU 1</th>
<th>BLEU 2</th>
<th>BLEU 3</th>
<th>BLEU 4</th>
<th>METEOR</th>
<th>ROUGE-L</th>
</tr>
</thead>
<tbody>
<tr>
<td>TextRank</td>
<td>2.55</td>
<td>1.74</td>
<td>1.02</td>
<td>0.56</td>
<td>1.63</td>
<td>2.24</td>
</tr>
<tr>
<td>RAKE</td>
<td>2.87</td>
<td>2.17</td>
<td>1.31</td>
<td>0.43</td>
<td>2.27</td>
<td>2.29</td>
</tr>
<tr>
<td>BERT_SENTENCE</td>
<td>2.83</td>
<td>1.99</td>
<td>1.26</td>
<td>0.67</td>
<td>2.21</td>
<td>2.73</td>
</tr>
<tr>
<td>BERT_PAGE</td>
<td>2.84</td>
<td>1.96</td>
<td>1.18</td>
<td>0.54</td>
<td>1.57</td>
<td>1.89</td>
</tr>
</tbody>
</table>

Surprisingly, the deep learning methods, BERT, do not outperform the traditional machine learning method, RAKE. It only performs the best when the metrics evaluate longer common sequences such as BLEU 4 and ROUGE-L. The possible reason is the data being analyzed in this study. One advantage of using deep learning models is the ability to analyze semantic meanings of the sentences. However, most of the content in the slides are sentences extracted from different paragraphs in papers or textbooks, and the meaning of text in each page is very different from another, which means the contextual meaning in sentences is lacking. Therefore, it is hard to take the advantage of BERT when the learning materials consist of phrases or incomplete sentences. We therefore suggest to represent the input in full text to leverage BERT. For instance, a textbook or papers.

For Research question 2, Does the quality of students’ highlights affect their learning performance?

We adopted RAKE as the text summarization method for our study and used the summaries as key concepts to evaluate the quality of students’ highlights. BLEU 1 is used to calculate the similarity between summaries and highlights since most of the highlights contain words or phrases and the length is often less than 2. The sum of highlight scores of 11 weeks is regarded as the quality level of highlights. Figure 2 shows the correlation between highlight score and students’ learning performance. The Spearman correlation is 0.75 with P-value less than 0.001, which indicates that the highlight score is highly correlated to learning performance. We assigned students into two groups. The top 20% students belong to HIGH_PERFORMANCE and others belong to LOW_PERFORMANCE. Since both groups show the normal distribution, a one-way ANOVA test is performed. Table 2 shows that the mean of high learning performance group is larger than low learning performance group at a statistical significance level.

![Figure 2](image.png)  
*Figure 2. The correlation between highlight score and students’ learning performance. The Spearman*
correlation is 0.75, p < 0.001.

Table 2. The one-way ANOVA test for group with high learning performance and group with low learning performance.

<table>
<thead>
<tr>
<th></th>
<th>Mean HP</th>
<th>SD HP</th>
<th>Mean LP</th>
<th>SD LP</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIGHLIGHT_SCORE</td>
<td>25.32</td>
<td>10.33</td>
<td>4.79</td>
<td>8.33</td>
<td>10.64**</td>
</tr>
</tbody>
</table>

***: p < 0.001, **: p < 0.01, *: p < 0.05; HP=High Performance, LP=Low Performance

The average of highlight scores for both groups through 11 weeks are represented in Figure 3. The group with high learning performance consistently achieved a better highlight score than the group with low learning performance except for the last week since that week was for the final exam and no slides were uploaded. The first week was the introductory week in which the instructor introduced the basic concepts about the course. Both groups were able to achieve a good highlight score. It is observed that when the difficulty level of lectures increased as the week goes by, the highlight scores for both groups decreased. After week 4, both groups were more familiar with the courses and were able to highlight the key concepts. As both groups improved their highlight scores, the gap also enlarges. This indicates that the more concepts students learned, the more likely that students who gain a high highlight score can achieve a high learning performance. Therefore, we conclude that highlight quality strongly affects students’ learning performance, and the impact is more significant as the learning period increases.

Figure 3. The average highlight score through 11 weeks.

4. Conclusion

In this research, we want to find a method to augment humans for automatically extracting the important concepts from learning materials and grading students’ highlights since the accuracy of human grading may be affected when the number of highlights is too large and when graders have different knowledge levels. The summaries of four different text summarization techniques are compared with the gold standard answer from the instructor using BLEU 1, BLEU 2, BLEU 3, BLEU 4, METEOR, and ROUGE. The results show that when the content of slides consists of phrases or incomplete sentences, RAKE outperforms other techniques including deep learning algorithms. Then we use the summaries from RAKE as reference to calculate the highlight score using BLEU 1 considering most highlights contain only a few words. Finally, we explored whether highlight quality has influence on students’ learning performance. The Spearman correlation between highlight quality and learning performance is 0.75 with P-value less than 0.001, which indicates that highlight quality is highly correlated to learning performance. The students were assigned into two groups. Since both groups show a normal distribution, a one-way ANOVA test is performed. The results show that the mean of highlight scores in high learning performance group is higher than the low learning performance groups at a statistical significance level. We further found that as students learned more concepts, the difference in highlight
score between two groups increased, which means students who achieve high learning performance are more likely to identify the key concepts from lectures.

In future work, we will apply the deep learning extraction model, BERT, on lectures that consist of full text, and expect to achieve a decent performance. Also, we will integrate this model into BoolRoll, provide feedback to students with the key concepts highlighted by the model, and investigate model's effectiveness by measuring students’ learning performance.

Acknowledgements

This work was partly supported by JSPS Grant-in-Aid for Scientific Research (S)16H06304 and NEDO Special Innovation Program on AI and Big Data 18102059-0.

References

Lin, C. Y. (2004, June). Looking for a few good metrics: Automatic summarization evaluation-how many samples are enough?. In NTCIR.
How Does The Quality of Students’ Highlights Affect Their Learning Performance in e-Book Reading

Albert YANG*, Irene Y.L. CHENb, Brendan FLANAGANc & Hiroaki OGATAc

aGraduate School of Informatics, Kyoto University, Japan
bDepartment of Accounting, National Changhua University of Education, Taiwan
cAcademic Center for Computing and Media Studies, Kyoto University, Japan
*yang.ming.35e@st.kyoto-u.ac.jp

Abstract: Measuring a student’s highlight quality is a way to understand how well a student understands the learning materials. In general, students highlight the words or sentences they consider as important concepts during their learning process. An often way to measure students’ highlight quality is to let graders give a highlight score after reading their highlights. However, when graders do not reach agreement on the quality of highlights and give different scores due to different background knowledge, the influence of highlight quality on learning performance may decrease. In this study, we explore the marker function on BookRoll and propose another way to grade students’ highlight quality in terms of its similarity to the reference answer. Students can add highlights as markers when reading learning materials. The instructor will highlight the important concepts in the learning materials as a reference answer. Graders will use this reference answer to grade students’ highlights. The average of students’ quiz scores will be used to denote their learning performance. The results indicate that the highlight quality has a larger influence on students’ learning performance than that of highlight frequency when graders evaluate the highlights along with a reference answer. Next, we divided students into two groups and performed a one-way ANOVA test to evaluate the influence of highlight quality on learning performance, and found that students who achieve high learning performance are also likely to achieve a high highlight score. We further compared the highlight quality and highlight frequency to see which indicator is more correlated with learning performance and found that highlight quality outperforms highlight frequency.

Keywords: e-Book reading, learning analytics, learning performance

1. Introduction

Using highlights is considered as a cognitive learning behavior which influences reading comprehension positively (Van Horne et al., 2016). Nian et al. (2019) and Yin et al. (2019) analyzed the frequency of e-reader functions on BookRoll such as NEXT, PREV and MARKER used by students to observe whether these behaviors are related to learning performance. They found that students who used highlights frequently had a tendency to achieve better learning performance. Al-khazraji (2019) let learners use highlights to mark the sentences they think are important, and found that the use of highlights greatly affects the effectiveness of learning. Yufan et al. (2020) proposed that in addition to analyzing the frequency of highlights, the area of highlights may also be related to learning performance. However, without considering the content of highlights, it is still possible to overuse or misuse them and decrease learning performance.

To overcome this issue, we conducted an experiment in a university course to investigate the influence of students’ highlight quality on their learning performance.

RQ1: Is it better for graders to have a reference answer when grading students’ highlights?
RQ2: What is the correlation between students’ highlight quality and learning performance?
RQ3: Is highlight quality a better predictor than highlight frequency for predicting students’ learning performance?
2. Experiment

2.1 Data Collection

In this study, we analyzed students’ highlight records of a 12-week course in a university. A total of 22 slides of learning materials were uploaded to BookRoll. BookRoll is an e-Book reading system (Flanagan & Ogata, 2017) developed by Kyoto University. Students’ e-Book reading actions in BookRoll have been introduced in detail by (Ogata et al., 2015) and (Flanagan & Ogata, 2018). There were 44 students enrolled in this course. Before each class, the instructor uploaded the slides of the class to BookRoll for students to preview. Students were asked to highlight the words or sentences they think are important using the marker function on BookRoll. There are two types of markers they can use, red marker for important concepts and yellow marker for concepts they feel difficult to understand. The instructor gave two graders a reference answer. Graders graded the highlight to see whether students highlight the important concepts. Figure 1 is an example of marker function on BookRoll. Finally, Students were given a quiz every week to measure their understanding level. The average of quiz scores were used to denote their learning performance.

![Figure 1. An example of marker function on BookRoll.](image)

2.2 Grading Highlight Score with Reference Answer

Two graders independently graded students’ highlights. They first graded the highlights based on their background knowledge and understanding about the course. Then they used the highlights from the instructor as a reference answer and graded the highlights again. To know the level of agreement on the highlight scores from different graders, the Cohen-Kappa-Coefficient of their grading result were being calculated.

2.3 Highlight Quality vs. Highlight Frequency

Frequency of adding highlights is often discussed when exploring the correlation between highlight function and students’ learning performance. In this study, highlight quality is measured by the average of highlight scores grade by humans, whereas highlight frequency is denoted as the number of highlights students add. Finally, learning performance is measured by the average of quiz scores. We compared highlight quality with highlight frequency to decide which indicator has a better correlation with learning performance.

2.4 Procedure

To answer our three research questions, we used Cohen-Kappa-Coefficient to compare two different grading approaches and evaluated which of it better reflects students’ learning performance. After that, a one-way ANOVA test is conducted to observe the impact of highlight quality on learning
performance. Finally, we measured whether highlight quality is a better indicator than highlight frequency for predicting learning performance. The result and discussion for each research question is listed as the followings.

3. Results and Discussion

For Research question 1, Is it better for graders to have a reference answer when grading students’ highlights?

Table 1 and Table 2 show the result of highlight scores graded by two graders with a reference answer and without a reference answer. The diagonal indicates that two graders gave the same score. The sum of diagonal scores (25) in Table 1 is larger than the sum of diagonal scores (17) in Table 2, which means the grading result is more consistent between two graders when a reference answer is given. Table 3 shows the Cohen-Kappa-Coefficient and the mean absolute error of two highlight scores, with reference answer and without reference answer, and the Spearman correlation between two grading approaches and students’ learning performance. When graders grade with their own knowledge level, the agreement level is only 0.179, which means the possibility of the agreement is probably occurring by chance. On the other hand, the agreement level of highlight score graded with a reference answer is satisfactory with 0.423, which indicates that the possibility of the agreement does not occur by chance. The mean absolute error (MAE) between two graders was calculated for both approaches to measure the average difference between graders. The MAE between highlight scores is 6.61 when a reference answer is provided and 10.84 when a reference answer is not provided. The result indicates that the difference between highlight scores is larger on average when a reference answer is not provided. The Spearman correlation between learning performance and highlight score with a reference answer is 0.779, which is also better than the result of 0.525 without a reference answer. From this result, we know that when graders follow a reference instead of relying on their own background knowledge to grade, the highlight score is more consistent and correlates better to the learning performance. Table 4 lists students’ highlight score, highlight frequency and their learning performance. The highlight score is computed by the average of two graders’ scores based on the reference answer.

Table 1. The result of highlight score based on reference answer.

<table>
<thead>
<tr>
<th>Grader2</th>
<th>Highlight Score</th>
<th>&gt;=90</th>
<th>&gt;=80</th>
<th>&gt;=70</th>
<th>&lt;70</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grader1</td>
<td>&gt;=90</td>
<td>11</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>&gt;=80</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>&gt;=70</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>&lt;70</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 2. The result of highlight score without reference answer.

<table>
<thead>
<tr>
<th>Grader2</th>
<th>Highlight Score</th>
<th>&gt;=90</th>
<th>&gt;=80</th>
<th>&gt;=70</th>
<th>&lt;70</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grader1</td>
<td>&gt;=90</td>
<td>8</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>&gt;=80</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>&gt;=70</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>&lt;70</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 3. The Cohen-Kappa-Coefficient for two grading approaches and the Spearman correlation between two grading approaches and the learning performance.

<table>
<thead>
<tr>
<th></th>
<th>With Reference Answer</th>
<th>Without Reference Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohen-Kappa-Coefficient</td>
<td>0.423</td>
<td>0.179</td>
</tr>
<tr>
<td>Mean Absolute Error</td>
<td>6.613</td>
<td>10.84</td>
</tr>
<tr>
<td>Spearman Correlation</td>
<td>0.779***</td>
<td>0.525***</td>
</tr>
</tbody>
</table>
Table 4. Students’ highlight score, highlight frequency and learning performance.

<table>
<thead>
<tr>
<th>Student</th>
<th>HS</th>
<th>F</th>
<th>P</th>
<th>Student</th>
<th>HS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>81.5</td>
<td>524</td>
<td>84</td>
<td>S23</td>
<td>72.5</td>
<td>376</td>
<td>94</td>
</tr>
<tr>
<td>S2</td>
<td>60</td>
<td>333</td>
<td>64</td>
<td>S24</td>
<td>65</td>
<td>333</td>
<td>75</td>
</tr>
<tr>
<td>S3</td>
<td>60</td>
<td>434</td>
<td>66</td>
<td>S25</td>
<td>73</td>
<td>439</td>
<td>72</td>
</tr>
<tr>
<td>S4</td>
<td>67</td>
<td>392</td>
<td>77</td>
<td>S26</td>
<td>71</td>
<td>328</td>
<td>70</td>
</tr>
<tr>
<td>S5</td>
<td>67</td>
<td>381</td>
<td>74</td>
<td>S27</td>
<td>76</td>
<td>323</td>
<td>60</td>
</tr>
<tr>
<td>S6</td>
<td>67</td>
<td>498</td>
<td>73</td>
<td>S28</td>
<td>62</td>
<td>344</td>
<td>78</td>
</tr>
<tr>
<td>S7</td>
<td>84.5</td>
<td>466</td>
<td>81</td>
<td>S29</td>
<td>88</td>
<td>365</td>
<td>86</td>
</tr>
<tr>
<td>S8</td>
<td>90</td>
<td>339</td>
<td>81</td>
<td>S30</td>
<td>90</td>
<td>323</td>
<td>89</td>
</tr>
<tr>
<td>S9</td>
<td>89.5</td>
<td>360</td>
<td>89</td>
<td>S31</td>
<td>88</td>
<td>333</td>
<td>92</td>
</tr>
<tr>
<td>S10</td>
<td>76.5</td>
<td>349</td>
<td>80</td>
<td>S32</td>
<td>85</td>
<td>318</td>
<td>84</td>
</tr>
<tr>
<td>S11</td>
<td>63.5</td>
<td>360</td>
<td>70</td>
<td>S33</td>
<td>98</td>
<td>397</td>
<td>95</td>
</tr>
<tr>
<td>S12</td>
<td>60</td>
<td>318</td>
<td>62</td>
<td>S34</td>
<td>99</td>
<td>333</td>
<td>90</td>
</tr>
<tr>
<td>S13</td>
<td>60</td>
<td>323</td>
<td>60</td>
<td>S35</td>
<td>87</td>
<td>530</td>
<td>97</td>
</tr>
<tr>
<td>S14</td>
<td>86</td>
<td>471</td>
<td>88</td>
<td>S36</td>
<td>79.5</td>
<td>318</td>
<td>77</td>
</tr>
<tr>
<td>S15</td>
<td>89.5</td>
<td>392</td>
<td>82</td>
<td>S37</td>
<td>65</td>
<td>318</td>
<td>81</td>
</tr>
<tr>
<td>S16</td>
<td>97</td>
<td>530</td>
<td>100</td>
<td>S38</td>
<td>78.5</td>
<td>344</td>
<td>83</td>
</tr>
<tr>
<td>S17</td>
<td>93.5</td>
<td>402</td>
<td>86</td>
<td>S39</td>
<td>86</td>
<td>434</td>
<td>78</td>
</tr>
<tr>
<td>S18</td>
<td>91</td>
<td>439</td>
<td>95</td>
<td>S40</td>
<td>63.5</td>
<td>355</td>
<td>77</td>
</tr>
<tr>
<td>S19</td>
<td>100</td>
<td>450</td>
<td>97</td>
<td>S41</td>
<td>93.5</td>
<td>344</td>
<td>85</td>
</tr>
<tr>
<td>S20</td>
<td>79</td>
<td>376</td>
<td>80</td>
<td>S42</td>
<td>89</td>
<td>386</td>
<td>100</td>
</tr>
<tr>
<td>S21</td>
<td>91</td>
<td>450</td>
<td>90</td>
<td>S43</td>
<td>78</td>
<td>530</td>
<td>65</td>
</tr>
<tr>
<td>S22</td>
<td>82.5</td>
<td>434</td>
<td>100</td>
<td>S44</td>
<td>74</td>
<td>318</td>
<td>60</td>
</tr>
</tbody>
</table>

HS=Highlight Score, F=Frequency, P=Learning Performance

For Research question 2, What is the correlation between students’ highlight quality and learning performance?

After deciding the approach to grade highlights, we want to know the influence of highlight quality on learning performance. We classified the students into two groups based on their learning performance. Students with top 20% learning performance were assigned to HIGH_PERFORMANCE, and the rest of students were assigned to LOW_PERFORMANCE. Figure 2 is a boxplot of the highlight score for two groups. It demonstrates that most students in HIGH_PERFORMANCE have a better highlight score than students in LOW_PERFORMANCE. Since both groups represent the normal distribution, a one-way ANOVA test is conducted to observe whether highlight score influences students’ learning performance. Table 5 indicates that the mean of highlight score in HIGH_PERFORMANCE is larger than LOW_PERFORMANCE at a statistical significance level. The result shows that students with high learning performance are likely to achieve a better highlight score. The plausible explanation is that they know what they know and what they do not know, and they are capable of highlighting the sentences or words that represent the important concepts in learning.
For Research question 3, Is highlight quality a better predictor than highlight frequency for predicting students’ learning performance?

Next, we want to know the correlation between the frequency of adding highlights and the highlight score, and whether the quality of highlights is more related to students’ learning performance than the frequency of adding highlights. Table 6 shows the Spearman correlation between highlight score, frequency of adding highlights and learning performance. The correlation between highlight score and highlight frequency is only 0.31, which means students who frequently add highlights may not achieve a high highlight score if they missed the important concepts or marked the concepts that are considered as less important. The possible reason is that they are not sure about what they know or what they do not know, which leads to the misuse of highlights. The result also shows that the score of highlights is more correlated with learning performance than that of frequency. This suggests that students who are able to grasp the important concepts in learning materials are more likely to achieve higher learning performance.

Table 6. The Spearman correlation between highlight score, highlight frequency and students’ learning performance.

<table>
<thead>
<tr>
<th></th>
<th>HS</th>
<th>HF</th>
<th>LP</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HF</td>
<td>0.31*</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>LP</td>
<td>0.77***</td>
<td>0.37*</td>
<td>1</td>
</tr>
</tbody>
</table>

***: p < 0.001, **: p < 0.01, *: p < 0.05; HS=Highlight Score, HF=Highlight Frequency, LP=Learning Performance

4. Conclusion

In this research, we conducted an experiment on two highlight grading approaches to investigate whether grading based on a reference answer is better than grading based on graders’ own knowledge level. The result shows that agreement on score is hard to achieve for graders without a reference answer. After finding the better grading approach, we performed a one-way ANOVA test to examine whether highlight quality has influence on students’ learning performance. The results demonstrated
that students who achieve high learning performance are more likely to achieve a high highlight score because they can highlight the important concepts and identify concepts they are not familiar with. Conversely, learners who are not able to get high learning performance also cannot achieve a high highlight score. They are likely to mark the concepts that are not important since they are not sure what they know and what they do not know. Finally, we measure the correlation among highlight quality, highlight frequency and learning performance. We calculated the Spearman correlation for three descriptive data. The correlation between highlight quality and highlight frequency is low, which means students who mark frequently are not guaranteed to achieve a high highlight score because they are not able to identify important concepts from the learning materials and misuse the highlight function. Also, the correlation between learning performance and highlight quality is higher than the correlation between learning performance and highlight frequency, which indicates that highlight quality is a better predictor than highlight frequency for predicting learning performance.

In future work, we want to explore more methods which leverage machine learning algorithms for grading highlights. For instance, a semi-automatic way in which the instructor provides a reference answer and students’ highlights are automatically graded using machine learning algorithms, or a full-automatic approach that applies machine learning algorithms to summarize the text as important concepts and the highlights can be automatically graded.

Acknowledgements

This work was partly supported by JSPS Grant-in-Aid for Scientific Research (S)16H06304 and NEDO Special Innovation Program on AI and Big Data 18102059-0.

References

Yufan, X., Xuewang, G., Li, C., Satomi, H., Yuta, T., Hiroaki, O., & Yamada, M. Can the Area marked in eBook Readers Specify Learning Performance. Companion Proceedings 10th International Conference on Learning Analytics & Knowledge (LAK20) (pp. 638-648)
An educational system with functions of guidance and adaptive advice to support problem solving based on basic concepts of statistics

Kanako SUZUKI a*, Tatsuhiro KONISHI b
aGraduate School of Integrated Science and Technology, Shizuoka University, Japan
bFaculty of Informatics, Shizuoka University, Japan
*suzuki.kanako.16@shizuoka.ac.jp

Abstract: Recently, statistics is a necessary subject not only for experts but for common people as well. In the field of education on statistics, it is required to present that improve learners’ motivation and promote deep understanding. We thought that problems using statistical concepts on the real world can satisfy such a condition. In this study, we have developed an educational system equipped with a guidance function that asks learners such questions and organizes data, and an adaptive advice function that gives advice according to the type of the wrong answer. Learners tackle the problem using our system and system-independent statistical tools. In addition, we conducted an evaluation experiment to evaluate the usefulness of the system. The results showed that the guidance function and the advisory function of this system are effective for improving the exercise achievement rate for the learner to proceed with learning, and that the learner feels that the acquired knowledge is useful.

Keywords: Statistics, Guidance, Adaptive advice, Learning support system

1. Introduction

In modern society, statistics is a subject required by many other than the experts. However, in the field of education, mainly rote learning is employed. There have been various discussions on ways for learners to improve their statistical skills. It is argued that it is important to learn not only statistical calculation process but also total problem solving process using statistical skills (Wild & Pfannkuch, 1999). Wild & Pfannkuchh propose the PPDAC cycle as the total process of statistical problem solving. In the PPDAC cycle, the problem solving process consists of the following 5 steps: Problem, Plan, Data, Analysis, and Conclusion. Smith also points out the importance of total problem solving process (Smith, 1998). He recommends to incorporate active learning strategies in order to develop students' statistical inference ability. The active learning strategy is to let learners carry out activities such as study design, data collection, and result analysis. From the above, we design exercises that learners solve more than just the statistical calculation process. Here, since exercises of the study design or the data collection must be relatively long-term ones, we design our system to handle exercises of the result analysis process that can be finished within one lesson.

It is difficult to say that the teachers give encouraging lessons to the learners to develop motivation and a deep understanding of statistics. In this study, we define exercises which can encourage learners’ motivation and deepen understanding of concepts in real world using statistics and data of the real world (e.g., Let’s find the most important property of the team in order to increase winning rate by analyzing the stats data of the 12 Japan professional baseball
teams in 2015!). We developed a learning support system which acts some of the teacher’s roles in such exercises. We suppose that there are learners who don’t know how to solve exercises and find an erroneous conclusion. These learners need individual guidance and adaptive advice. However, teachers generally do not have enough resources to pay attention to individual needs because there are too many learners per teacher. There is also a limitation on school hours. In this study, we define a teacher’s role as conducting guidance and giving the adaptive advice to complete the exercise. We developed a learning support system which has the following two functions to help learners:

1. Guidance of learner's behavior toward completion of exercises.
2. Evaluation and adaptive advice of answers obtained from learners.

We evaluated our system experimentally and found that it works effectively.

There are many existing learning support systems for statistics. For example, the Web-based statistical learning support system (Baharun & Porter, 2012) aims to support mathematical problem solving rather than real world problems that our system supports. A learning support system by an expert system (Barbara & William, 1996) deals with real-world problems. However, the system does not have the functions for resolving learners’ impasse (the function of guiding the problem-solving process or the function of giving advice according to the error of the learner) that our system equips.

2. Methods

2.1 Discussion on Exercises Employed in the Learning Support System

Exercises in this learning support system are the exercises to find properties of a dataset using concepts of statistics. Learners complete the exercises using a statistical analysis tool (We adapt BellCurve for Excel in this study. It is statistical analysis software that adds statistical analysis functions to the Excel menu, such as calculating correlation coefficient, drawing box plot or scatter plot and so on).

In this study, as mentioned in Section 1, our system supports exercises to find properties of real world by using statistics in order to increase learners’ motivation and to deepen understanding. We define “deep understanding” as the understanding of how to use a statistical concept in the real world. We think that it is not deep understanding to know a statistical concept as a simple numerical formula. It tends to be hard for learners to understand how to use the statistical concept in real life because learners recognize only the mathematical aspect of the statistical concept as performed in in-class exercises. Therefore, our learning support system consists of situations from daily experience that learners can solve using the statistical concept. Our educational support system gives learners guidance after asking to execute such examples. It focuses on the method of statistical concepts using which learners solve the exercises. Our learning support system sets exercises that learners can use in real life. Learners solve these exercises with a statistical concept called the correlation coefficient.

2.2 Methods Guiding Learners for Solving Exercises

It is our idea, learners have abilities to solve exercises by themselves through organizing data and deciding which statistical concept to use. However, it is difficult to say that all learners can solve exercises by themselves. Therefore, our learning support system guides learners who cannot solve exercises by themselves. It shows what to do next in a given exercise to achieve a predefined goal. First, we defined the common learning processes independent of learning content.

(Step 1) Our learning support system gives a rough guide to learners for solving exercises.
(Step 2) Learners visualize the given data and information using the statistical analysis tool.
(Step 3) Learners solve sub-exercises to organize data from our learning support system.
(Step 4) Learners answer given exercises based on organized information in Step 3.

At first, our learning support system gives a rough guide to learners for solving exercises (Step 1). An example of an exercise for learning about correlation coefficient is shown below.
Let us compare the properties of the team with a high winning rate from the team results of the 12 professional baseball teams of Japan in 2015 and compare it with that of 1950! Learners should organize given data and visualize them using BellCurve for Excel first. Our system gives learners instructions on how to use BellCurve for Excel because in order to get the output correlation matrices and heat maps of correlation matrixes (Step 2).

In Step 3, our learning support system guides the learners by setting sub-exercises (Figure 1). The sub-exercises let learners focus on a parameter or correlation between parameters that is useful for solving the exercise. Then it gives learners adaptive advice according to mistakes that they make. Our system repeats setting sub-exercise in order to guide learners to achieve the goal of the exercise (Figure 1). Finally, it set a kind of another sub-exercise that has the same goal as the final goal of the exercise. We call such sub-exercise "final sub-exercise". Figure 1 shows example of dialogue setting sub exercise (in Step 3). The messages in Figure 1 are translated from Japanese into English for this paper.

2.3 Generation Method of Adaptive Advice According to Mistakes

Our learning support system checks the learners’ answers and gives learners adaptive advice according to their mistakes. Our system gives advices not only to suggest their answer is wrong simply, but also to show how they should correct their wrong answer. It prevents them from finding answer by trial-and-error. Variations of the correlation coefficient advice implemented in our system are shown below.

- If learners focus on the wrong cell, then point out that it’s wrong and that they should focus on another cell.
- If learners answer suggests that they misunderstand how to interpret positive/negative correlation coefficient, then let learners confirm meaning of positive/negative correlation coefficient by reading the textbook of statistics again.
- If learners answer suggests that they misunderstand how to interpret of strength correlation coefficient, then let learners confirm meaning of strength of correlation coefficient by reading the textbook of statistics again.

And variations of the box plot advice implemented in our system are shown below.

- If learners focus on the wrong box plot, then point out that it’s wrong and that they should focus on another box plot.
- If learners answered a number in wrong cell instead of another cell that has correct answer, then point out that you are focusing on wrong cell and read the textbook of statistics again if you miss some concepts of statistics.

First, we explain how to generate a method of adaptive advice according to mistake. Our learning support system has candidate for answers to all sub-exercises Learners select one of the
candidates to answer the sub-exercises. Our system has functions to diagnose learners’ answers. The learners are allowed to go to the next sub-exercises when his/her answer is correct. In case that the answer is wrong, our educational system gives adaptive advice according to his/her mistake. We divide all choices into correct answers and several types of typical mistakes. We attach adaptive advice to each typical mistake. The advice is designed to help learners fix their wrong answers. The above process is as shown in Figure 2. The messages in Figure 2 are also translated from Japanese into English for this paper. We attach adaptive advice to each typical mistake.

![Figure 2](image-url)  
**Figure 2. Example of advice according to mistake**

### 3. Design of our System

Our system is based on the assumption that the learner’s personal computer has a statistical analysis tool independent of the system. We independently developed our system and statistical analysis tool because by incorporating a statistical analysis tool in the system, the system itself has the advantage of being able to monitor the behavior of the learner, but we thought that it would be very costly to change the statistical analysis tool. For the learners to proceed smoothly with the exercise, we created slides explaining how to use and explaining the statistical terms used in the exercise. Learners refer to these slides in addition to our system and to proceed with learning.

We mainly devised the message display when implementing our system. The points that we have devised specifically are as follows.

- Admixed words of praise according to the behavior of the learners
- Directives and commentary output in a colloquial style

### 4. Experimental Evaluation of the Experiment

#### 4.1 Hypothesis

In order to evaluate the effectiveness and usefulness of the system, we conducted experimental evaluation on the following two hypotheses:

**Hypothesis 1:** Guidance and advisory functions are useful for learners to proceed with learning.

**Hypothesis 1-1:** The guidance function is effective.
Hypothesis 1-2: The advisory function is effective.
Hypothesis 2: Subjective evaluation by the subjects suggest our system is useful
  Hypothesis 2-1: The guidance function of this system is useful for learners. The subjects feel
  the guidance function is useful.
  Hypothesis 2-2: The subjects feel the advisory function is useful. Hypothesis
  Hypothesis 2-3: The subjects feel our system totally useful.

4.2 Outline of the Experiment

First, we describe the design of the experiment. The subjects are ten persons (six master's students, one
university student, two university graduates and one vocational school graduate). To verify hypotheses
1-1 and 1-2, we compare the incidence of impasses with and without guidance and advisory functions.
Specifically, the subject is made to solve the problem under the following three conditions, and the
accuracy rate of the final sub-exercises is compared (final sub-exercise is defined in 3.2).
  (Condition 1) Solve exercises using a system without guidance and advisory functions.
  (Condition 2) Solve exercises using a system equipped with only guidance function.
  (Condition 3) Solve exercises using a system equipped with both guidance and advisory functions. In
  order to verify hypotheses 2-1, 2-2 and 2-3 a questionnaire was given to the subjects after the
  experiment (Condition 3). The questions in the questionnaire are as fellows.
   (Q1) Was the guidance function useful?
   (Q2) Was the advisory function useful?
   (Q3) Was the system equipped with both guidance function and advisory function useful?

The subjects select one from 5 (useful), 4 (relatively useful), 3 (fair), 2 (relatively useless) and
1 (useless). There are six exercises to be solved by the subject, which consist of three exercises solved
using the correlation coefficient and three exercises solved using the box plot. The difficulty level of
each exercise is different, with number 1 being the easiest and number 3 being the most difficult. We
let subjects solve each exercise under the above conditions (1)-(3) and compare the correct answer rate
of the final sub exercises. The time limit for the subjects to solve an exercise is seven minutes. Subjects
are allowed to give up solving exercises.

4.3 Result on Hypothesis 1

Table 1 shows the average of the correct answer rate of each exercise. Some final sub-exercises ask
multiple items. “equipped” means the condition that the system equipped the function written in the
topline of Table 1.

Table 1. Problem difficulty level / correct answer rate by system

<table>
<thead>
<tr>
<th>Themes</th>
<th>Exercise Level</th>
<th>Guidance function equipped</th>
<th>Advisory function equipped</th>
<th>Guidance function equipped</th>
<th>Advisory function equipped</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation Coefficient</td>
<td>Level1</td>
<td>85.00%</td>
<td>98.33%</td>
<td>98.33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Level2</td>
<td>30.00%</td>
<td>80.00%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Level3</td>
<td>95.00%</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Box Plot</td>
<td>Level1</td>
<td>83.33%</td>
<td>98.33%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Level2</td>
<td>86.66%</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Level3</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
</tbody>
</table>

In all of the exercises, the correct answer rate of the “condition with guidance/without advice”
is higher than (or equal to) that of the “condition without guidance/without advice”. This result suggests
that the guidance function was effective in resolving the impasses.
In all of the exercises, the correct answer rate of the “condition with guidance/without advice” is higher than (or equal to) that of the “condition without guidance/without advice”. This result suggests that the advisory function was effective in resolving impasses. As a result, Hypothesis 1 was supported.

4.4 Result on Hypothesis 2

Table 2 shows the distribution of responses and average values for the three questions in the questionnaire. The results of (Q1) and (Q2) indicate that all subjects have positive evaluation of both the guidance function and the advisory function. We interviewed subjects who answered “fair” to (Q3). They had high grade of problem-solving ability, so they correctly answered the final exercises without any advice. Therefore, they felt the system is not so effective. As a reason for subjects’ evaluation, they wrote the following messages (We translate the original messages in Japanese into English). “It was easier to carry out the exercises because I could find what I should do next and how my answer was wrong.”, “I think (Condition 3) is upwardly compatible with (Condition 2), but I felt the detailed messages from (Condition 3) a little bothersome when I could solve exercises without advice.”

<table>
<thead>
<tr>
<th>Question Number</th>
<th>Selection</th>
<th>useless (=1)</th>
<th>relatively useless (=2)</th>
<th>fair (=3)</th>
<th>relatively useful (=4)</th>
<th>useful (=5)</th>
<th>Average Point</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Q1)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>4.9</td>
<td></td>
</tr>
<tr>
<td>(Q2)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>4.9</td>
<td></td>
</tr>
<tr>
<td>(Q3)</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>4.4</td>
<td></td>
</tr>
</tbody>
</table>

This result suggests that most of the subjects feel the guidance and advisory functions of our system useful. Furthermore, it is suggested that the subjects feel that the system having both functions is useful for learning. As a result, Hypothesis 2 was supported.

5. Conclusion

In this study, we set out the exercise of finding out the property from the data by using the statistical concepts, we have developed a learning support system to act as a teacher. We developed two functions. One is a guidance function that promotes data organization. The other is an advisory function that gives adaptive advice according to learners’ typical mistakes. In addition, we evaluated the usability and educational effectiveness of our system experimentally. The results show that guidance and advisory functions increase the exercises achievement rate. Furthermore, the results of questionnaires on the guidance function and the advisory function show usefulness of our system. In the future, we will tackle the issues pointed out by the subjects. For example, introducing GUI to our system.

References


OpenLA: Library for Efficient E-book Log Analysis and Accelerating Learning Analytics

Ryusuke MURATA*, Tsubasa MINEMATSU & Atsushi SHIMADA

*Graduate School of Information Science and Electrical Engineering, Kyushu University, Japan
Faculty of Information Science and Electrical Engineering, Kyushu University, Japan

*murata@limu.ait.kyushu-u.ac.jp

Abstract: This paper introduces an open source library for e-Book (digital textbook) log analysis, called OpenLA. An e-Book system is a useful system which records learning logs. Various analysis using these logs have been conducted. Although there are many common processes in preprocessing logs, the functions have been developed by per researcher. To reduce such redundant development, OpenLA provides useful modules to load course information, to convert learning logs into a more sophisticated representation, to extract the required information, and to visualize the data. OpenLA is written in the Python language and compatible with other Python libraries for analysis. This paper provides a brief explanation of each module, followed by re-implementation samples of related studies using OpenLA. The details about OpenLA is open to public at https://www.leds.ait.kyushu-u.ac.jp/achievements.

Keywords: Python, learning analytics, e-Book, digital textbook

1. Introduction

Thanks to the widespread of information and communications technology (ICT) and digital learning systems, we can collect not only learning results, such as examination results, but also learning and studying processes of individuals, such as how much time a learner spends to study. Understanding learner behavior is crucial in learning analytics (LA). Learning logs collected via digital systems are often utilized for analytics for teaching and learning. A learning management system is a digital system that is generally used for collecting learning logs; however in recent years, e-Book (digital textbook) systems are increasingly used. An e-Book system records detailed learning processes, such as when a learner opens a learning material, turns a page in the material, highlights, notes, and bookmarks.

The e-Book operation logs are used for research, e.g., to determine the successful features in learning activities (Yin, 2019); to understand learner behaviors (Shimada, 2019), to estimate academic performance (Okubo, 2018), and to identify at-risk students (Shimada, 2018). The first step of such research is to aggregate the learning logs in order to extract learning behaviors, such as calculating the reading time of each learner and page-wise summary of operations by learner. So far, each researcher has developed his/her study’s preprocessing, even though there are many common processes. Individually developing the common processes causes redundancy and decreases efficiency of advanced learning analysis.

One of the solutions is to develop an open source library for the common processes. For example, the computer vision field has many common processes such as segmentation, calibration, and optical flow. However, redundancy is reduced by an open source library named OpenCV (https://opencv.org/). In addition, various open source libraries have been developed such as OpenGL (https://www.opengl.org/) for 3D computer graphics and PTAM (Klein, 2007) for augmented reality.

We developed such an open source library for e-Book log analysis, called “OpenLA.” This library reduces redundancy in the development of common processes and accelerates the development of core technologies for advanced learning analytics. In Section 2, we explain OpenLA’s application programming interfaces (APIs). In Section 3, we show usage examples of OpenLA. In Section 4, we describe how to activate OpenLA. Lastly, in Section 5, we conclude our paper and indicate areas for improvement.
2. API Concept

2.1 Basic Information

The APIs are written in Python language and compatible with other Python libraries for analysis, such as Scikit-learn (Pedregosa, 2011) and Tensorflow (Abadi, 2016). The dataset used in this library has the same structure with that of the open source ones used to conduct data challenge workshops in LAK19 and LAK20 (https://sites.google.com/view/lak20datachallenge). Note that the dataset is not a unique structure, and other e-Book systems can be used for constructing this dataset. The dataset includes four types of CSV files:

- **Course # EventStream.csv**: Data of the logged activity from learners’ interactions with the BookRoll system (Ogata, 2015).
- **Course # LectureMaterial.csv**: Information about the length of lecture materials used.
- **Course # LectureTime.csv**: Information about lecture schedules.
- **Course # QuizScore.csv**: Data on the final score of each student.

For analyzing this dataset, getting course information, converting the learning logs into a form suitable for analysis, extracting the required information, and visualizing the data are essential and common preprocessing for e-Book log analysis. To reduce redundant development, OpenLA provides four types of modules: Course Information, Data Conversion, Data Extraction, and Data Visualization. Figure 1 shows the flow of preprocessing with OpenLA. In the following section, we describe the four types of modules and data forms.

![Figure 1. The flow of preprocessing with OpenLA.](image)

2.2 Course Information Module

The Course Information module receives the dataset files and loads basic information about a course. This module returns the instance of Python class `CourseInformation`, and the member function returns basic information including registered user id, content id (lecture materials), users’ final score, and lecture start and end time. Table 1 shows a part of the member functions for loading basic information.

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>OpenLA.CourseInformation.load_eventstream()</td>
<td>Load the event stream data</td>
</tr>
<tr>
<td>OpenLA.CourseInformation.user_ids()</td>
<td>Get the user ids in this course</td>
</tr>
<tr>
<td>OpenLA.CourseInformation.lecture_start_time()</td>
<td>Get the lecture start time in this course</td>
</tr>
</tbody>
</table>

Table 1.

Part of Member Functions of the CourseInformation class
The function `load_eventstream` in this module loads the event stream (learning logs) as a Pandas.DataFrame type member variable of the Python class named `EventStream`. Table 2 shows an example of event stream data. The class `EventStream` has useful member functions to aggregate data in an event stream. However, detailed information cannot be aggregated from the original event stream. Therefore, an event stream needs to be converted into a more sophisticated representation by the Data Conversion module, and the required information must be extracted by the Data Extraction module.

<table>
<thead>
<tr>
<th>user id</th>
<th>contents id</th>
<th>operation name</th>
<th>page no.</th>
<th>event time</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>X</td>
<td>OPEN</td>
<td>1</td>
<td>2020/01/01 12:00</td>
</tr>
<tr>
<td>A</td>
<td>X</td>
<td>NEXT</td>
<td>1</td>
<td>2020/01/01 12:05</td>
</tr>
<tr>
<td>A</td>
<td>X</td>
<td>MARKER</td>
<td>2</td>
<td>2020/01/01 12:12</td>
</tr>
</tbody>
</table>

Table 2.

An Example of an EventStream

2.3 The Data Conversion Module

The Data Conversion module provides four functions; `convert_into_operation_count`, `convert_into_page_wise`, `convert_into_page_transition`, and `convert_into_time_range`. Each function receives an instance of `EventStream` and converts it into more sophisticated Python classes named `OperationCount`, `Page-wiseAggregation`, `PageTransition`, and `Time-rangeAggregation`. As with `EventStream`, each Python class has the converted log data as their member variable, and they also have the member functions to aggregate the data. These four classes are used depending on the kind of analysis conducted. Each class is described below.

Firstly, `OperationCount` provides the total number for each operation in each content. Table 3 shows an example data of `OperationCount`. The tendencies in students’ learning activities can be understood from this content-level aggregation.

<table>
<thead>
<tr>
<th>user id</th>
<th>contents id</th>
<th>NEXT</th>
<th>PREV</th>
<th>MARKER</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>X</td>
<td>75</td>
<td>32</td>
<td>19</td>
</tr>
<tr>
<td>A</td>
<td>Y</td>
<td>169</td>
<td>106</td>
<td>11</td>
</tr>
<tr>
<td>B</td>
<td>X</td>
<td>60</td>
<td>18</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 3.

An Example of OperationCount

Secondly, `Page-wiseAggregation` provides data on how long each user reads a page and how many times each operation is used on the page. Table 4 shows an example data of `Page-wiseAggregation`. This representation is useful to investigate which pages are intensively read by high-performance students, on which pages learners take notes and draw highlights, and so on.

<table>
<thead>
<tr>
<th>User id</th>
<th>Contents id</th>
<th>page no.</th>
<th>reading seconds</th>
<th>MARKER</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>X</td>
<td>1</td>
<td>109</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>X</td>
<td>2</td>
<td>245</td>
<td>3</td>
</tr>
<tr>
<td>A</td>
<td>X</td>
<td>3</td>
<td>195</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4.

An Example of Page-wiseAggregation
Thirdly, PageTransition also provides data on how long each user reads each page and how many times each operation is used. However, it also considers the reading order (going back and jumping page), whereas Page-wiseAggregation provides the total value for each page. Table 5 shows an example data of PageTransition. This representation accurately tracks learning activity; therefore, it is suitable for analyzing reading behavior patterns.

Table 5. An Example of PageTransition

<table>
<thead>
<tr>
<th>user id</th>
<th>contents id</th>
<th>page no</th>
<th>reading seconds</th>
<th>time of entry</th>
<th>time of exit</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>X</td>
<td>1</td>
<td>60</td>
<td>2020/01/01 12:00</td>
<td>2020/01/01 12:01</td>
</tr>
<tr>
<td>A</td>
<td>X</td>
<td>2</td>
<td>245</td>
<td>2020/01/01 12:01</td>
<td>2020/01/01 12:05</td>
</tr>
<tr>
<td>A</td>
<td>X</td>
<td>1</td>
<td>49</td>
<td>2020/01/01 12:05</td>
<td>2020/01/01 12:06</td>
</tr>
</tbody>
</table>

Finally, Time-rangeAggregation provides the pages read for the longest time and the number of operations used in each time range. Table 6 shows an example data of Time-rangeAggregation. The period is specified in the converting function (60 seconds is specified in Table 6). This representation is useful for comparing user activity within a period, such as whether a student’s reading behavior is following the reading behavior of a teacher or other students.

Table 6. An Example of Time-range Aggregation

<table>
<thead>
<tr>
<th>user id</th>
<th>contents id</th>
<th>elapsed seconds</th>
<th>page no</th>
<th>start of range</th>
<th>end of range</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>X</td>
<td>0</td>
<td>1</td>
<td>2020/01/01 12:00</td>
<td>2020/01/01 12:01</td>
</tr>
<tr>
<td>A</td>
<td>X</td>
<td>60</td>
<td>4</td>
<td>2020/01/01 12:01</td>
<td>2020/01/01 12:02</td>
</tr>
<tr>
<td>A</td>
<td>X</td>
<td>120</td>
<td>5</td>
<td>2020/01/01 12:02</td>
<td>2020/01/01 12:03</td>
</tr>
</tbody>
</table>

2.4 The Data Extraction Module

The Data Extraction module receives an instance of EventStream or converted representations and extracts the required information on users, specific contents, and so on. For example, to analyze which learning behavior affects student performance, one may focus on the logs of high-score students and low-score students. To investigate the logs of these two types of students, one needs to extract their logs. OpenLA provide two functions for this extraction: users_in_selected_score and select_user. Table 7 shows the functions of the Data Extraction module.

Table 7. Part of the Functions of the Data Extraction Module

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>OpenLA.select_user()</td>
<td>Extract data about selected user</td>
</tr>
<tr>
<td>OpenLA.select_contents()</td>
<td>Extract data about selected content</td>
</tr>
<tr>
<td>OpenLA.select_by_lecture_time()</td>
<td>Extract data during/ before/ after the lecture</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
2.5 The Data Visualization Module

Data Visualization module receives an instance of EventStream or converted representations, and renders a visual graph. Data features of the graph can be easily understood, and user data is easy to compare. Table 8 shows the functions of the Data Visualization module. For example, the function `visualize_pages_in_time_range` renders a line graph that shows which pages are read by users in a given time period. The graph shows which pages a student is interested in, and makes it easy to compare the differences among students’ activities.

Table 8.
Part of the Functions of Data Visualization Module

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>OpenLA.visualize_time_series_graph()</td>
<td>Visualize learning activity in a time series</td>
</tr>
<tr>
<td>OpenLA.visualize_operation_count_bar()</td>
<td>Visualize operation count of a specific user</td>
</tr>
<tr>
<td>OpenLA.visualize_pages_in_time_range()</td>
<td>Visualize the page user read in a time range</td>
</tr>
</tbody>
</table>

3. Usage Example

To show how effective OpenLA is in reducing redundant development and writing short code, we adapt OpenLA for off-task detection in lecture time (Akçapınar, 2019). This research aims to detect off-task behaviors from in-class reading activity. Students took a lecture with an open-book quiz during the last 15 minutes; therefore, users’ reading patterns during this time varied. The researchers eliminated the quiz part, and the remaining learning logs were grouped into one-minute intervals. After the grouping, they extracted pages read for each student at each time interval. The page difference between a student and teacher at each time interval represents the student’s reading pattern vector. The clustering result of the vectors identified on-task or off-task students.

For preprocessing, we grouped student reading behavior in one-minute intervals and visualized the results. Figure 2 shows summary of preprocessing, the number of lines without OpenLA, and the corresponding functions of OpenLA. The whole code and the result is shown in documentation website (https://www.leds.ait.kyushu-u.ac.jp/achievements). To aggregate learning logs at each interval, we utilized the function, `convert_into_time_range`, which converts EventStream into Time-rangeAggregation data. The arguments of `convert_into_time_range` were specified, considering that the in-class activity log is grouped into one-minute intervals and the last 15 minutes is eliminated. After the conversion, the function `visualize_pages_in_time_range` in the Data Visualization module visualized students’ reading tracks. The functions can also be applied to preprocessing for other time-based analysis.

As shown in this example, OpenLA reduces the workload of preprocessing. The functions contribute not only to reduce the amount of codes, but also to give clear definition, i.e., what is the input and output of each function. Although this example shows the efficient preprocessing with OpenLA, making dataset files for this library is additional work; therefore, we will implement the function to make dataset files from a database.

Figure 2. Summary of preprocessing for off-task detection.
4. Installation of OpenLA

The supported version is Python 3.7.X, and the required libraries are pandas 0.25.X, numpy 1.16.X, and matplotlib 3.1.X. To install OpenLA, one has to type “pip install OpneLA” into the command line in your Python environment. The required libraries are installed together if not previously installed. Note that PIP must be installed (https://pypi.org/project/pip/). The documentation of OpenLA and a sample dataset are open to public (https://www.leds.ait.kyushu-u.ac.jp/achievements).

5. Conclusion and Future Works

OpenLA is a Python library and provides four useful modules to preprocess e-Book learning logs. The Course Information module loads basic information of a course and loads Event Stream. The Data Conversion module converts the event stream into a more sophisticated representation. The Data Extraction module extracts the required information. The Data Visualization module visualizes the data. These modules help to reduce the redundant development of common preprocessing so that researchers can focus on the development of core technologies for advanced learning analytics.

In future work, we will receive feedback from end users and improve performance and usability. Moreover, we will improve datasets requirement. The current version of OpenLA requires four types of CSV files as described in Section 2.1; therefore, end users need to export the CSV files from a database on their own. In addition, analyzing online (real time) logs with OpenLA is difficult, because OpenLA requires CSV files. To solve these problems, we plan to add modules to connect with databases.

Acknowledgements

This work was supported by JST AIP Grant Number JPMJCR19U1, and JSPS KAKENHI Grand Number JP18H04125, Japan

References


Sequence Pattern Mining for the Identification of Reading Behavior based on SQ3R Reading Strategy

Owen H.T. Lua, Anna Y.Q. Huang, Che-Yu Kuo, Irene Y.L. Chen, Stephen J.H. Yang

College of Computer Science, National Pingtung University, Pingtung City, Taiwan
Computer Science & Information Engineering, National Central University, Taoyuan City, Taiwan
Department of Accounting, National Changhua University of Education, Changhua City, Taiwan
anna.yuqing@gmail.com

Abstract: SQ3R (survey, question, read, recite, and review) is an efficient reading method that has confirmed the benefits of learning performance in numerous studies. To demonstrate that the e-book reading environment can also obtain the benefits of SQ3R, we conducted a course with 60 students, classify students into two groups based on the SQ3R ability. The results show that in the context of e-book reading, students' behavior: (1) create a memo and modify it regarding the reviewing content (CMeP) and (2) create a memo and modify it at the next login (CMeO) is related to recite and review steps in SQ3R. Furthermore, the above two e-book reading behaviors are positively related to students' learning performance and learning engagement.

Keywords: SQ3R, reading strategy, computer-assisted learning

1. Introduction

With the rapid development of learning technology, computer-assisted learning has gradually spread to schools at all levels. In recent years, the advantage of the digital learning environment built by computer technology lies in the ability to record students' historical learning actions in the learning log. Therefore, by measuring the students' engagement from learning logs, students' learning situation can be effectively and deeply to explore. Therefore, measuring the student's engagement through the learning log has attracted the attention of some researchers (Flanagan & Ogata, 2017; Jovanović et al., 2017; Yin et al., 2017). In addition, the use of learning techniques is also a key factor for effective learning. Therefore, the introduction of e-books into learning has become a new trend in recent years. From the above description, this study has introduced ebook reading environment named as BookRoll (Ogata et al., 2015; Flanagan & Ogata, 2017) into course Then, this study aims to extract actions from students’ tracking logs in learning environment to explore students’ learning patterns.

In the learning process, reading is the basic foundation of learning. Besides, the reading strategies adopted by students will have an impact on the reading performance. Therefore, this study will explore students’ reading strategies in e-books. To guide students to perform efficient reading, Robinson (1946) has proposed SQ3R reading strategy which consists of the five steps of Survey, Question, Read, Recite, Review. Therefore, the research questions in this study are proposed as following.

- **RQ1**: In reading e-book process, exploring the relationship between reading actions and learning outcome.
- **RQ2**: For the groups classified by SQ3R reading strategy, explore the relationship between SQ3R reading strategy and learning performance.
2. Methods and experiments

2.1 Participates

The experiment was conducted with 60 students, with a total of eight weeks at a university in Taiwan. The course was divided into before-class reading material preview, classroom learning and after-class review. Classroom learning includes teacher lectures, concept summaries, exercise exercises and quizzes, and after-class review includes after-class reflection and homework exercises. In this paper, we focus on students’ activities on before-class preview. Students’ reading log will be recorded by BookRoll (Ogata et al., 2015, Flanagan & Ogata, 2017).

2.2 Learning actions in BookRoll

Students’ participation in this course can be observed through the log of the BookRoll system. According to the proposed reading functionalities in BookRoll, this study encoding students’ reading actions from the logs. We define nineteen codes from the logs recorded by the students for each operation of the e-book, and divide them into four categories: Page, Bookmark, Marker, Memo. Page is a page-related action such as Next, Prev, Jump. Bookmark is a bookmark-related action, such as Add Bookmark and Delete Bookmark. Marker is a tag-related action, such as Add Marker, Delete Marker and Marker. Memo is related to the memo-related action such as Add Memo, Delete Memo, Change Memo and Memo. Table 1 shows the constructed 19 codes from BookRoll.

This study defines three types of reflection process, which are the same page, jump page, and re-login. Take Change Memo as an example, Same-page modification (CMe) means that when a student modifies memo, he has stayed on the same page since he last modified or added this memo. Page-change modification (CMeP) means that when a student modifies memo, he has changed the paged since he last modified or added this memo. Re-registration (CMeO), which means that when a student modifies memo, he has re-logged in the system since he last modified or added this memo. The original Change Memo will be recoded into the above three categories as CMe, CMeP, and CMeO. The other three actions (Delete Marker, Delete Memo, and Delete Bookmark) are also divided into three categories according to the same logic.

Table 1. The codes of reading actions in BookRoll.

<table>
<thead>
<tr>
<th>Category</th>
<th>Code</th>
<th>Description</th>
<th>Code Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Page</td>
<td>N</td>
<td>Go to page forward</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>Go to page backward</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>Jump to a specific page</td>
<td>J</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>Open the ebook</td>
<td>O</td>
</tr>
<tr>
<td>Marker</td>
<td>AMa</td>
<td>Add a marker to the current page</td>
<td>AMa</td>
</tr>
<tr>
<td></td>
<td>DMa</td>
<td>Adding a marker to the current page and then deleting this marker during viewing the current page</td>
<td>AMa…DMa</td>
</tr>
<tr>
<td></td>
<td>DMaP</td>
<td>Add a marker to this page, and come back to delete this marker after viewing the others pages</td>
<td>AMa…[N/P/J]…DMa</td>
</tr>
<tr>
<td></td>
<td>DMaO</td>
<td>Add a marker to this page, and come back to delete this marker after re-login this ebook</td>
<td>AMa…[O]…DMa</td>
</tr>
<tr>
<td>Memo</td>
<td>AMe</td>
<td>Add a memo to the current page</td>
<td>AMe</td>
</tr>
<tr>
<td></td>
<td>DMe</td>
<td>Adding a memo to the current page and then deleting this memo during viewing the current page</td>
<td>AMe…DMe</td>
</tr>
<tr>
<td></td>
<td>DMeP</td>
<td>Add a memo to this page, and come back to delete this memo after viewing the others pages</td>
<td>AMe…[N/P/J]…DMe</td>
</tr>
<tr>
<td></td>
<td>DMeO</td>
<td>Add a memo to this page, and come back to delete this memo after re-login this ebook</td>
<td>AMe…[O]…DMe</td>
</tr>
<tr>
<td></td>
<td>CMe</td>
<td>Adding a memo and then change the content in this memo during viewing the current page</td>
<td>AMe…CMe</td>
</tr>
</tbody>
</table>
### 2.3 SQ3R learning strategies for ebook reading actions

SQ3R reading strategy aims to guide students how to perform effectiveness reading to improve their reading comprehension, and SQ3R reading strategy which include of Survey, Question, Read, Recite, and Review steps is proposed by Robinson (1946). SQ3R emphasizes browsing the full text structure first, and then reading in detail. It is considered to be a very effective reading strategy, which helps to understand the full text and increase the memory retention rate. The survey step aims to grasp the structure of the article and the logic of the author to arrange the structure of the article, have a preliminary understanding of the full text. The question step is mainly to ask questions as a guide for the next stage of reading. According to the previous questions, read step focus on finding the answer from the reading article, and write down the answer in students’ own words. The question proposed by students themselves can be guide students reading article into more directional to guide students more understanding and improve students’ retention memory of article. In addition, students’ question can also make teacher to know the level of students’ reading comprehension. After finishing reading article, recite step aims to guide students to summary and recite the content for the key points of article by themselves. Students can back to read article again when they forget the content of key points. Without open the textbook, review step focus on guiding students to recall the article’s structure and raised question to answer. If student can't answer the raised question, go back to the recite step.

### 3. Results and discussion

To reply RQ1 (In reading e-book process, exploring the relationship between reading actions and learning outcome), this study applied Spearman correlation analysis to explore the relationship between the actions and learning performance. The Spearman correlation coefficient is a test statistic based on the covariance to measure the statistical relationship or association between two variables. For representing students’ reading actions, some researchers have constructed ebook reading actions (Yamada et al., 2017). Based on the extracted ebook reading actions from previous studies. Table 1 showed the constructed 19 actions from learning logs in this study. Table 2 shows the descriptive statistics results for the extracted actions and Spearman correlation coefficient.

#### Table 2. The results of Spearman correlation between the reading codes and learning outcome.

<table>
<thead>
<tr>
<th>Category</th>
<th>Actions</th>
<th>Mean/Std.</th>
<th>Spearman Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Page</td>
<td>N</td>
<td>1591.67/878.59</td>
<td>.11</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>711.75/524.3</td>
<td>.09</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>37.72/45.55</td>
<td>.21</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>114.15/53.61</td>
<td>.10</td>
</tr>
<tr>
<td>Marker</td>
<td>AMa</td>
<td>145.23/111.28</td>
<td>.29*</td>
</tr>
<tr>
<td></td>
<td>DMa</td>
<td>9.73/11.67</td>
<td>.33*</td>
</tr>
<tr>
<td></td>
<td>DMaP</td>
<td>1.62/2.7</td>
<td>.29*</td>
</tr>
</tbody>
</table>

...
This study has extracted actions include of Page, Marker, Memo, and Bookmark categories. From the Spearman correlation results showed in Table 2, the actions belong to the Marker and Memo categories have significant relationship with learning performance. The range of Spearman correlation coefficient values for Marker and Memo categories are .29~.33, and .27~.45, respectively. It means that the actions belong to Marker and Memo categories have positive relationship with learning performance.

To reply RQ2 (For the groups classified by SQ3R reading strategy, explore the relationship between SQ3R reading strategy and learning performance), this study firstly classified students through their reading actions, and then extracted reading patterns through LSA. From the previous research studies, we can know that the SQ3R reading strategy is one of the most effective reading strategy. In SQ3R reading strategy, the question, recite, and review steps in the SQ3R strategy has great influence on students’ reading comprehensive. Corresponding to reading actions in BookRoll, the AMe, CMe, CMeP, CMeO actions are the most related to question, recite, and review steps in the SQ3R. Therefore, this study will apply the k-means method to classify students' reading by using AMe, CMe, CMeP, CMeO actions. We classified students’ reading actions into high (G_H) and low (G_L) reading engagement groups. For the G_H and G_L groups, Table 3 shows that the descriptive results of AMe, CMe, CMeP, CMeO actions.

<table>
<thead>
<tr>
<th>Category</th>
<th>Actions</th>
<th>Mean/Std.</th>
<th>Spearman Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMaO</td>
<td>3.83/6.86</td>
<td>.33**</td>
<td></td>
</tr>
<tr>
<td>Memo</td>
<td>AMe</td>
<td>53.6/32.73</td>
<td>.45***</td>
</tr>
<tr>
<td></td>
<td>DMe</td>
<td>0.67/1.16</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td>DMeP</td>
<td>0.68/1.32</td>
<td>.16</td>
</tr>
<tr>
<td></td>
<td>DMeO</td>
<td>0.25/0.57</td>
<td>.42**</td>
</tr>
<tr>
<td></td>
<td>CMe</td>
<td>18.05/14.87</td>
<td>.41**</td>
</tr>
<tr>
<td></td>
<td>CMeP</td>
<td>31.07/28.96</td>
<td>.26*</td>
</tr>
<tr>
<td></td>
<td>CMeO</td>
<td>15.0/22.58</td>
<td>.27*</td>
</tr>
<tr>
<td>Bookmark</td>
<td>AB</td>
<td>3.03/5.9</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>DB</td>
<td>0.93/1.24</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>DBP</td>
<td>0.25/0.79</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>DBO</td>
<td>0.13/0.59</td>
<td>-.05</td>
</tr>
</tbody>
</table>

From Table 3, the number of G_H and G_L groups are 19 and 41, respectively. Besides, the students’ midterm score in G_H group has significantly higher than the students’ midterm score in G_L group (t=2.96, p<.01). Besides, the reading engagement of AMe (t=9.26, p<.001), CMe (t=9.23, p<.001), CMeP (t=4.54, p<.001), and CMeO (t=5.36, p<.001) for students in G_H group are significantly higher than students in G_L group. This is the reasons for students in G_H group can obtain significantly higher midterm score than students in G_L group.

### 4. Conclusion

This study aims to demonstrate the SQ3R benefits in the e-book. We extract 31 reading actions from the BookRoll log, and 9 reading actions belonging to the marker and memo categories are positively correlated with the students' learning performance. The result shows that students' learning performance is influence by the students' number of reading actions. On the other hand, to prove the assumption, this study classifies students into two groups based on the SQ3R reading strategy. We can observe that students with high engagement will perform the recite and review steps of SQ3R through reading.
patterns related to CMeP and CMeO patterns. Furthermore, based on the reading patterns extracted from the LSA, this study considered five association rules which could enable students to improve their reading engagement. According to the results, students with a higher number of reading patterns related to CMeP and CMeO actions will have a higher probability of higher engagement.

Acknowledgments

This work is supported by Ministry of Science and Technology, Taiwan under grants MOST-109-2511-H-008-007-MY3, MOST-108-2511-H-008-009-MY3, MOST-106-2511-S-008-004-MY3, and Ministry of Education, Taiwan.

References


Improving Skill for Self-Reviewing Presentation with Robot

Ryotaro SEYA* & Akihiro KASHIHARA
The University of Electro Communications, Japan
1-5-1 Chofugaoka, Chofu, Tokyo 182-8585
*ryotaro.seya@uec.ac.jp

Abstract: Presentation is very important for researchers to present their work. Proper presentation requires them to review it before actual presentation, which contributes to refining their presentation. In this work, we have focused on self-review, in which researchers review their presentation by themselves. A general way of self-review is to make a video of presentation, and then check it out. However, it brings about uncomfortable feeling due to looks and voice on the video, which prevents self-reviewing. We have accordingly developed a robot as presentation avatar, which reproduces presentation novel researchers (as learners) make. On the other hand, we have also confirmed that learners often overlook points to be modified in their nonverbal behavior reproduced by the robot. The main issue addressed in this paper is how to improve self-review skill with presentation robot. Our approach to this issue is to diagnose learners’ presentation by means of the presentation behavior model, and to help them compare their self-review results with the diagnosed results. Such comparison allows them to promote awareness of unsuitable/insufficient points in their self-review, which contributes to self-review skill development. In this paper, we demonstrate robot presentation comparison with two robots, in which one robot reconstructs their presentation with self-review results and the other one reconstructs it with diagnosed results. This paper also reports a case study with the system whose purpose was to ascertain whether robot presentation comparison contributes to promoting awareness of unsuitability/insufficiency in self-review. The results suggest the possibility of self-review skill development.

Keywords: Presentation, Nonverbal Behavior, Self-Review, Skill Improvement, Diagnosis, Robot

1. Introduction

Presentation is one of important activities for researchers to present their work. Proper presentation requires them to review it before actual presentation, which contributes to refining their presentation. Since it is particularly important to perform nonverbal behavior such as gesture and paralanguage according to presentation intention (Goto et al., 2018), it is essential to review nonverbal behavior in addition to presentation slides.

There are two types of reviews, which are self-review and peer review. In self-review, researchers review their presentation by themselves. In peer review, researchers could obtain reviews of their presentations from their peers including more skillful lab members. Conducting self-review before peer review allows researchers to compare between self-review results and peer-review ones, and to become aware of points to be modified that could not be unaware in self-review, which contributes to improving their self-review skill.

We have been addressing the issue how to help novice researchers as learners self-review their presentation to become aware of points to be modified in their presentation (Inazawa & Kashihara, 2017, 2018). As a general way of self-review, learners could make a video of their presentation, and then check it out. However, they would feel quite uncomfortable due to their looks and voice on the video. This uncomfortableness prevents them from self-reviewing (Inazawa & Kashihara, 2017). We have accordingly developed a robot as presentation avatar, which reproduces presentation learners make. We have also designed a model of presentation behavior to provide a checklist including points to be reviewed, whose purpose is to help learners check if their nonverbal behavior is conducted as
intended during presentation reproduced by the robot. The model illustrates how presentation intentions could be accomplished with nonverbal behavior. The results of the case study suggest that the robot avatar can significantly reduce uncomfortable feeling in self-review, and that the checklist facilitates awareness of points to be modified (Keisuke Inazawa & Akihiro Kashihara, 2017, 2018).

On the other hand, we have also confirmed that learners often overlook points to be modified in their nonverbal behavior even if they are provided with the checklist. In order to resolve this problem, it is necessary to improve their self-review skill so that they can become aware of unsuitability and insufficiency in self-review, and obtain new perspectives in self-review.

The main issue addressed in this paper is how to improve self-review skill with presentation robot. Our approach to this issue is to diagnose nonverbal behavior in presentation by learners with the presentation behavior model, and to help them compare their self-review results with the diagnosed results. In this paper, we demonstrate a self-review system with two robots, in which one robot reconstructs their presentation with self-review results and the other one reconstructs it with diagnosed results. They could compare nonverbal behavior conducted by the two robots to become aware of unsuitability and insufficiency of their self-review.

This paper also reports a case study with the system whose purpose was to evaluate the self-review improvement with two robots. The results suggest the possibility that it promotes awareness of unsuitable and insufficient points in self-review.

2. Self-Review System

2.1 Framework

Figure 1 shows the framework of the self-review system for promoting awareness of unsuitability and insufficiency in self-review. It has three phases, which are recording, diagnosis, and presentation comparison by means of two robots. In phase 1, the system requires learners to make presentation with their presentation documents (P-documents for short) including slides. It also captures the presentation data including P-documents, motion, and voice by means of PC and Microsoft Kinect (Microsoft Corporation, 2018).

In phase 2, the system diagnoses the presentation (the captured data) by following the presentation behavior model shown in Figure 2 (Inazawa & Kashihara, 2018), which we have designed by referring to related work on presentation for research (Nancy, 2008; Karia, 2014; Kurihara et al., 2007; Zhao et al., 2015; Chollet et al., 2015; McNeill, 1992; Kamide et al., 2014; Melinger & Levelt, 2004). In presentation, it is necessary to achieve presentation intentions such as attention awakening and attention control by means of presentation behavior. The model accordingly represents the correspondence of presentation intentions to nonverbal behavior for achieving them. In the model, there are three layers, which are presentation intentions, behavior category (corresponding to detailed

![Figure 1. Framework for Improving Self-Review Skill.](image-url)
intentions), and components of presentation behavior. Each behavior category represents nonverbal behavior achieving its corresponding intention, which consist of a number of its components. Using the model, the system determines whether learners’ nonverbal behavior is appropriate for achieving their presentation intentions.

On the other hand, it is hard to infer the intentions from the captured data. Before phase 2, the system accordingly requires learners to use the checklist shown in Figure 3 to enter their intentions in presenting each slide. Using the captured data and entered intentions, the system attempts to extract unsuitable/insufficient nonverbal behavior. We currently restrict presentation intentions to detailed ones related to “attention to contents”. We also restrict nonverbal behavior to the one obtained from the corresponding components involving gestures (face direction, deictic gesture), paralanguage (voice volume, pitch), and text decoration associated with the gestures. In the self-review, learners are accordingly expected to use the checklist to tick the review points as to face direction, deictic pointing, paralanguage, and text decoration.

In phase 3, the system first requires learners to self-review using the robot system and checklist we have developed so far (Inazawa & Kashihara, 2018). In this self-review process, the robot reproduces the recorded presentation, and expects them to check if their nonverbal behavior is suitable/sufficient for achieving their presentation intentions entered. If they find out unsuitable/insufficient points, they are expected to tick them in the checklist. We call this initial self-review. Second, the system uses two robots to help learners think the differences between their initial self-review results and the diagnosed results via comparison of presentation conducted by the two robots.

This phase is composed of two steps. In step 1, one robot reproduces presentation learners make in phase 1, and the other robot reconstructs their presentation with their initial self-review results. The system expects learners to reflect on their nonverbal behavior reconstructed to confirm the validity of their initial self-review. In step 2, one robot reconstructs their presentation with the initial self-review.

<table>
<thead>
<tr>
<th>Presentation Intention</th>
<th>Behavior Category</th>
<th>Components</th>
<th>Review Points</th>
<th>Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attention to oral/slide contents</td>
<td>Attention Control</td>
<td>Face Direction</td>
<td>Directing face to slide</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Deictic Pointing</td>
<td>Pointing to slide contents</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Attention Awakening</td>
<td>Paralanguage</td>
<td>Emphasizing important points by volume/pitch</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Text Decoration</td>
<td>Emphasizing important points by color/bold/underline</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. Checklist.
results, and the other one reconstructs it with the diagnosed results. The system also expects them to become aware of unsuitability and insufficiency in their initial self-review.

2.2 Presentation Diagnosis

Using captured data including gesture (face direction and deictic pointing), oral explanation (contents of speech and paralanguage), and slide contents, the system recognizes nonverbal behavior. As for gesture, the system classifies the face direction into three according to face angle recognized by Kinect: screen direction, audience direction and looking down. The system also identifies the pointing position on the screen from the pointing gesture recognized by Kinect. The screen is divided into nine regions. The pointing position is identified as one of the regions.

As for oral explanation, the system converts the oral contents to text using NTT TechnoCross Voice-Series as speech recognition engine. It also obtains the paralanguage using Praat that is an application software extracting volume/pitch of voice during explanation.

In addition, the system obtains text/figures/tables and their positions on each slide from P-document. It also extracts the information of text decoration, which are used to diagnose nonverbal behavior. In entering presentation intentions on slides, learners are required to attach the labels of attention control/attention awakening to text/figures/tables to which they want to attract attention.

The recognized nonverbal behavior and entered intentions are stored with the time sequence as presentation scenario. Using this scenario, the system then diagnoses the unsuitable/insufficient points. The system decides whether gestures/paralanguage/text decoration are appropriate for attracting attention to text/figures/tables embedding the presentation intentions.

2.3 Robot Presentation Comparison

In robot presentation comparison, the system uses Sota (Vstone Co., Ltd., 2017) produced by Vstone as presentation robots. The Two Sotas reproduce/reconstruct presentation conducted by learners, and expects them to find out unsuitable/insufficient points in their self-review. Learners watch presentations conducted by two robots. They could also pause and redo these presentations during watching.

In reconstructing presentation with initial self-review results, the system replaces nonverbal behavior learners self-reviewed as inappropriate with the one appropriate for achieving the presentation intention, which allows the robot to conduct the reconstructed presentation. On the other hand, there are cases where nonverbal behavior learners self-reviewed as inappropriate is diagnosed as appropriate by the system. Such self-review called over-review is not considered in presentation reconstruction. In reconstructing presentation with diagnosed results, in addition, the system replaces inappropriate nonverbal behavior in the same way.

In demonstrating reproduced presentation, Sota executes the captured nonverbal behavior according to its time sequence. However, Sota reproduces captured oral explanation with higher pitch than original, whose purpose is to reduce uncomfortableness in self-review. In demonstrating reconstructed presentation, Sota also executes the presentation scenario that stores recognized nonverbal behavior and its time sequence. Sota faces forward (to audience) to explain with voice having the same speed as learners, but changes its face direction to the slide and points to the corresponding region during explaining the text/figure/table embedding presentation intention. Sota also changes the voice with higher pitch and more volume.

3. Case Study

3.1 Preparation and Procedure

We had a case study whose purpose was to ascertain whether robot presentation comparison could be beneficial for promoting awareness of unsuitability/insufficiency in self-review. Comparing initial self-review results and self-review results confirmed/modified from the robot presentation comparison, we ascertained the benefits of the self-review system.
We conducted this study in the following procedure with 8 participants who were graduate and undergraduate students in informatics and engineering. First, we explained to them how presentation should be conducted with nonverbal behavior according to the presentation behavior model. The participants were then required to prepare their presentation with a P-document we provided in advance. In the presentation preparation, they could decorate texts in the P-document. The participants were also required to enter presentation intentions (attention control and attention awakening).

The presentation was recorded. After a 10-minute interval, they were required to use the checklist to self-review their presentation reproduced by one robot (Initial self-review). After a 10-minute interval, they were twice required to compare presentations conducted by the two robots, and to self-review with checklist. In the first step, they compared their presentation reproduced and the one reconstructed with the initial self-review results (Second self-review). In the second step, they also compared the reconstructed presentation and the presentation reconstructed with the results diagnosed by the system (Third self-review). After each comparison, the participants were required to answer several questions.

To evaluate to what extent the participants could become aware of points to be modified as to nonverbal behavior for achieving their presentation intentions, we calculated the recall ratios of modified points found in the initial, second, and third self-review. Recall ratios were calculated as the ones of the number of modified points found by the participants to the number of modified points diagnosed by the system. The hypotheses we set up in this study were as follows:

H1: The first comparison allows the participants to confirm their initial self-review,
H2: The second comparison allows them to become aware of unsuitable/insufficient points in their initial self-review, and
H3: The self-review system allows them to improve their initial self-review.

3.2 Results

Figure 4 shows the average recall ratios in each self-review. The p-value corresponding to the F-statistic of one-way ANOVA is higher than 0.10 which suggests there was no significant difference between these self-reviews. However, Figure 4 shows a decrease in recall ratios from the initial self-review to second self-review. This suggests that the first comparison of robot presentation promotes not confirming but reconsidering the initial self-review for the worse. This does not support H1.

As shown in Figure 4, there was an increase in recall ratios from the second self-review to third self-review. This suggests that the second comparison promotes awareness of unsuitable/insufficient points in the initial self-review, which supports H2. In addition, Figure 4 shows an increase in recall ratios from the initial self-review to third self-review. This tends to support H3.
4. Conclusion

In this paper, we have proposed a self-review system that allows novice researchers to become aware of unsuitable and insufficient points in self-reviewing nonverbal presentation behavior with two robots. The self-review system diagnoses nonverbal behavior in learners' presentation with the presentation behavior model, and helps them think the differences between their initial self-review results and the diagnosed ones via the robot presentation comparison, in which one robot reconstructs their presentation with initial self-review results and the other one reconstructs it with diagnosed results.

From the results of the case study, the robot presentation comparison could promote awareness of unsuitable and inappropriate points in self-review, which contributes to improving self-review skill. On the other hand, some participants could not take notice of differences between nonverbal behavior conducted by two robots, which suggests the necessity of more active support in robot behavior comparison.

In addition to such active support, in future, we will address the issue how to exaggerate nonverbal behavior, with which the differences between nonverbal behavior conducted by two robots could become more obvious.

Acknowledgements

The work is supported in part by JSPS KAKENHI Grant Number JP18K19836 and JP20H04294.

References


Middle-School Students’ Behavior Pattern and Strategy Selection in Problem Solving: A Study Based on Data from PISA 2012

Yang LIU\textsuperscript{a}, Bo YANG\textsuperscript{a}, Lan WU\textsuperscript{b}, Baoping LI\textsuperscript{ab} & Shengquan YU\textsuperscript{ab}
\textsuperscript{a}Advanced Innovation Center for Future Education, Beijing Normal University, China
\textsuperscript{b}Faculty of Education, Beijing Normal University, China
*libp@bnu.edu.cn

Abstract: Vary-one-thing-at-a-time (VOTAT) strategy is regarded as the optimal strategy in the knowledge acquisition stage of complex problem solving (CPS) in many studies. Based on the log-file data of the Climate Control task of computer-based assessment of the CPS in the 2012 cycle of the Programme for International Student Assessment (PISA), 388,931 pieces of records from 20,597 students who used VOTAT were collected for an in-depth analysis. The result of the latent class analysis identified three kinds of strategies, the depth-first strategy, breadth-first strategy and mixed-strategy. The Pearson’s chi-squared Test and Welch variance analysis showed that the task performance of the students in the three groups varied. These results suggest that VOTAT strategy could be distinguished in a more elaborated way, which would contribute to portray the CPS process in detail, and so it is proved to play a key role in the CPS performance again.

Keywords: complex problem solving (CPS), learning analytics, VOTAT, depth-first strategy, breadth-first strategy, mixed-strategy

1. Introduction

As countries around the world are attaching great importance to the development of students’ domain-general skills. The competence of CPS, which is regarded as the key of the future education (OECD, 2013), has attracted more and more attention of educators.

During the process of complex problem solving, the application of the VOTAT strategy usually makes it easier to find the answers. Although a lot of researchers found that the use of the VOTAT strategy was significantly correlated with higher performance in problem-solving tasks, most of them focused on a simple control action of variables, the continuity between behaviors was not taken into account, which would limit the understanding of CPS.

2. Literature Review

2.1 Definition and Assessment of Complex Problem Solving

Complex problem solving (CPS), also known as dynamic problem solving, dynamic decision making, interactive problem solving, and creative problem solving in PISA 2012 (Greiff, Molnár, Martin, Zimmermann, & Csapó, 2018), can be seen as finding solutions in dynamic tasks. There is insufficient information to solve the problem at the given state and the problem solver needs to integrate the information obtained in the process of exploration to solve the problem (Buchner, 1995). In terms of structure, CPS is mainly divided into two dimensions: knowledge acquisition and knowledge application (Funke, 2001).

2.2 VOTAT Strategy in CPS
Tschirgi (1980) proposed the VOTAT strategy, which is also known as the control of variables strategy in the manipulation-of-variables task (Wüstenberg, Greiff, & Funke, 2012; Croker & Buchanan, 2011; Kuhn & Dean, 2005). At the stage of knowledge acquisition, VOTAT represents the fact that the problem solver explores the independent influence of input variables on output variables by changing a single input variable. The empirical research showed that the use of the VOTAT strategy was significantly correlated with higher performance of complex problem solving (Wüstenberg, Stadler, Hautamäki, & Greiff, 2014; Greiff, Wüstenberg, & Avvisati, 2015).

However, Molnár et al. (2018) found that the use of the VOTAT strategy did not always lead to higher performance and only conscious users of the VOTAT strategy proved to be the best solvers with non-conscious users of the VOTAT strategy the second and non-VOTAT strategy users the worst. Therefore, based on the common VOTAT strategy, this paper aims to make a more detailed exploration to further understand the cognitive pattern in the process of CPS, and increases the interpretability and educational significance of the strategy.

3. Research Questions

In this work, we aim to answer the following questions:
- Are there multiple strategies linked to VOTAT in the exploration of problem solving?
- What are the differences in the performance of students using different strategies in both the specific task and overall problem-solving proficiency?

4. Research Design

4.1 Test Task and Data

This paper opted for using the data (retrieved from https://www.oecd.org/pisa/pisaproducts/database-chapisa2012.htm) obtained from the CPS test unit, Climate Control in PISA 2012 as it provided the authoritative log-file of VOTAT. There are about 30,820 students from 42 countries or economies participated. Only the process data of manipulating input variables with a total of 477,258 entities was collected. In the data, some students did not use VOTAT in the knowledge acquisition phase and their data was deleted and finally, 388,931 pieces of records from 20,597 students were available for subsequent analysis.

![Climate Control](image.png)

**Figure 1.** The Test Task of Climate Control in PISA 2012
4.2 Variables and Scoring

4.2.1 Coding the Behavior in the Climate Control Task

All the behaviors of participants were recorded in the log-file. In this dynamic interaction system, information would not be generated in the final drawing behavior. Therefore, only manipulating data in the process of exploration such as clicking the “reset” tab and adjust the control button will be coded and analyzed.

The subsequent behavior of VOTAT was counted and labeled by whether a certain behavior occurred and the behavior which occurred was labeled with 1 while the behavior which did not occur was labeled with 0. There are four main types of behaviors:

- The previous VOTAT behavior was repeated and no input variable values were changed. This kind of behavior sequence was labeled as “e (0_step_0_class_0_value)”.  
- Based on the previous input variable settings, only different values were tried changed on the same control variable. This kind of behavior sequence was labeled as “ev (1_step_0_class_1_value)”. 
- Based on the previous input variable settings, only one of the other variables was changed. This kind of behavior sequence was labeled as “ec (1_step_1_class_0_value).”  
- Based on the previous input variable settings, a new input variable is manipulated and the previous control variable was restored to the initial value.  This kind of behavior sequence was labeled as “ecv+ (2_step_1_class_1_value)”. 

Other behaviors are not used in this study because their counts tend to zero-inflate. Table 1 shows some of the behavior sequences.

Table 1. VOTAT and Its Follow-up Behaviors

<table>
<thead>
<tr>
<th></th>
<th>[2,0]_0</th>
<th>[1,0]_0</th>
<th>[0,1]_0</th>
<th>[0,0]_1</th>
<th>[1,0]_1</th>
<th>[2,0]_1</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1,0]_0</td>
<td>ev</td>
<td>e</td>
<td>ecv+</td>
<td>ecv+</td>
<td>ec</td>
<td>---</td>
</tr>
<tr>
<td>[0,1]_0</td>
<td>ecv+</td>
<td>ecv+</td>
<td>e</td>
<td>ecv+</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>[0,0]_1</td>
<td>ecv+</td>
<td>ecv+</td>
<td>e</td>
<td>ecv+</td>
<td>e</td>
<td>ec</td>
</tr>
</tbody>
</table>

Note: The column represents the VOTAT behaviors and the row represents the exploration behaviors after the VOTAT behaviors. The three values in brackets show the input variable values of the top control, central control and bottom control in the Climate Control task, respectively.

4.2.2 Students’ Performance in Problem-Solving Tasks

In the Climate Control task, scores would be awarded if the correct relationship between the input and output variables was given. In addition to this task which involves interactive behaviors of the students in the process of problem solving, some other tasks such as logical reasoning and information retrieval were also employed in PISA 2012 to assess students’ problem-solving ability and an overall score was given. For the overall problem-solving proficiency, PISA used an imputation methodology to derive 5 plausible values, the plausible values were random draws from the marginal posterior of the latent distribution for each student (OECD, 2014). This paper used the first plausible value as the overall score of the students (cf., Greiff et al., 2015).

5. Results

5.1 Exploratory Latent Class Analysis of Various Follow-up Behaviors of VOTAT

The latent class analysis was conducted using Mplus. In this paper, three latent class models were obtained. Table 2 indicates the fit indices.

Table 2. Fit Indices for Latent Class Analyses
As Table 2 indicates, the 3-class model fits the data best since all of its fit indices appear smallest among the three models with the entropy of 0.946 in Table 2 implying that the 3-class model is reliable. In addition, each class is named according to the latent class probability and conditional probability of four main behaviors after VOTAT. As Figure 2 illustrates, the latent class 1 shows high probability in ev, which indicates that the students using this class preferred to explore different values of the same control variable after VOTAT. The latent class 2 shows the same probability as class 1 in e, ev and ec but a higher probability in ecv+, which indicates that the students using this class explored different values of the same control after VOTAT, as well as a single control of the others. In the latent class 3, the high probability of ec means different control variables were explored.

The depth-first search and breadth-first search are traversal algorithms to search graph structure and tree structure in computer science (Kreher & Stinson, 1999; Jungnickel, 2008). Depth-first search starts from the initial node, and extends the search to the next level of child node sequentially, which is similar to the behavior pattern of the latent class 1 in the previous analysis. Breadth-first search starts from the initial node and first cover the neighboring nodes, which is similar to the behavior pattern of the latent class 3. Therefore, the latent class 1, 2 and 3 are named depth-first strategy, mixed-strategy and breadth-first strategy, respectively. For example, the schematic diagram of a problem solver who used depth-first strategy as shown in Figure 3, and the code ev is marked with a red border.

5.2 Performance Differences of Using the Three Strategies in the Climate Control Task

Pearson’s chi-squared test was carried out to investigate the difference of three strategies in the performance of the Climate Control task, as Table 3 shows. The result indicated that there is a significant difference among the students who used different strategies ($\chi^2 = 2534.374, p = 0.000***$). The
correlation coefficient of 0.351 indicated a moderate association, which means the identification of three strategies is reliable (Rea & Parker, 2014).

Table 3. Pearson’s Chi-Squared Test for the Three Different Strategies in the Climate Control Task

<table>
<thead>
<tr>
<th></th>
<th>Depth-First</th>
<th>Breadth-First</th>
<th>Mixed-Strategy</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td>9170</td>
<td>432</td>
<td>4684</td>
<td>14286</td>
</tr>
<tr>
<td>Incorrect</td>
<td>3253</td>
<td>1627</td>
<td>1431</td>
<td>6311</td>
</tr>
<tr>
<td>Sum</td>
<td>12423</td>
<td>2059</td>
<td>6115</td>
<td>20597</td>
</tr>
</tbody>
</table>

\[ \chi^2 = 2534.374 \quad P = 0.000*** \quad \nu = 0.351 \]

It was found that students using the mixed-strategy performed the best with 77.112% of them correctly finishing the task. In comparison, students with breadth-first strategy did the worst with only 17.58% of them correctly completing the task. The performance of the students with the depth-first strategy also did a good job with 75.923% them succeeding. There is one thing worth noting that the percentage of the students correctly finishing the task was calculated based on the final sampling weight in PISA and therefore, it would not completely agree with the sample distribution in Table 3.

5.3 Differences of the Three Strategies in the Overall Problem-Solving Performance

Moreover, to explore students’ overall performance in CPS, the data was first analyzed by using the normality test. The Kolmogorov-Smirnov test revealed that the data did not follow the Gauss Distribution \(D = 0.016, p = 0.000***\). Then, the homogeneity of variances test showed the assumption also failed \(\text{Bartlett’s } K - \text{squared} = 61.426, p = 0.000***\). Therefore, the data could not be analyzed using the conventional ANOVA method and the Welch’s test was used. The result showed a moderate effect with the effect size of 0.069 \(\text{Welch } F = 848.805, p = 0.000***, 1 - \beta > 0.99\). Table 4 indicates the result of the subsequent Games-Howell’s multiple comparison.

Table 4. Games-Howell’s Multiple Comparisons of the Three Strategies in the Overall Problem-Solving Performance

<table>
<thead>
<tr>
<th>Multiple Comparisons</th>
<th>Mean Difference</th>
<th>Standard Error</th>
<th>Sig</th>
<th>95% confidence interval lower-bound</th>
<th>95% confidence interval upper-bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth-First Mixed-Strategy</td>
<td>-21.4249</td>
<td>1.287</td>
<td>0.000</td>
<td>-24.4416</td>
<td>-18.4083</td>
</tr>
<tr>
<td>Depth-First Breadth-First</td>
<td>61.8619</td>
<td>1.9064</td>
<td>0.000</td>
<td>57.3915</td>
<td>66.3323</td>
</tr>
<tr>
<td>Mixed-Strategy Breadth-First</td>
<td>83.2868</td>
<td>2.0213</td>
<td>0.000</td>
<td>78.5475</td>
<td>88.0262</td>
</tr>
</tbody>
</table>

The result complied with the preceding finding about students’ performance in the Climate Control task and individuals with mixed-strategy students performed the best, the depth-first strategy students the second and the breadth-first students the worst.

6. Discussion and Conclusion

This paper has identified three different types of strategies linked to VOTAT by using the latent class analysis to explore the follow-up behaviors of VOTAT, the depth-first strategy, breadth-first strategy, mixed-strategy. Similar findings were mentioned, such as engineering design (e.g., Ball, Evans, Dennis, & Ormerod, 1997) and information seeking (e.g., Heinström, 2005), but in the research of CPS, related research was limited (at the best of our knowledge). This research proved that CPS performance was strongly related with the strategies after applying VOTAT, which would be worth exploring.

The students with the mixed-strategy outperformed the other two kinds in terms of CPS performance, the depth-first strategy ranked the second and the last was the breadth-first strategy. A hypothesis may exist that the difference among the three groups may be related to the cognitive load. From the perspective of cognitive psychology, problem solving is defined as information-seeking in problem space from the initial state to the target state (Newell & Simon, 1979). As the breadth-first strategy involves the exploration in both variable types and values, the cognitive load is increasing rapidly with multiple attempts. However, the capacity of human information processing is limited.
(Miller, 1956) and this is why students with the breadth-first strategy would be more likely to fail. By contrast, the depth-first strategy only involves the attempts in different variable values, which delivered a low cognitive load and therefore possible better performance. Furthermore, the mixed-strategy group achieves a more efficient way of information integration with a lower level of cognitive load by focusing on only one variable and value in each step, which can be considered as a higher-level variant of the VOTAT strategy.

According to the previous findings, strategies play a crucial role in solving problems. Teachers could actively guide students in using the effective strategies and continuous training can be offered to help students improve metacognitive skills.

References


A warm-up for adaptive online learning environments – the Elo rating approach for assessing the cold start problem

Maciej PANKIEWICZ

Institute of Information Technology, Warsaw University of Life Sciences, Poland
maciej_pankiewicz@sggw.pl

Abstract: The aim of this study is to present and evaluate the Elo rating algorithm as a tool for assessing the task difficulty in terms of the so-called “cold-start” problem – during the initial phase of the introduction of the adaptive system to the public. This analysis has been performed on the real data originating from the online programming course available on the RunCode platform: the online learning environment with multiple attempts allowed and feedback provided after every attempt. There have been 50055 submissions on 76 tasks uploaded by 299 RunCode users. It has been found that the Elo rating algorithm achieves the correlation of 0.702 with the reference values already for the sample size of n = 5, and the correlation of 0.905 for the sample size of n = 50. The Elo algorithm outperforms the Proportion Correct method for small sample sizes and may be a more reasonable choice as a simple method for task difficulty estimation during the initial phase of introducing the adaptive system to the public.

Keywords: Difficulty estimation, Elo rating, proportion correct, automated assessment, programming course

1. Introduction

The problem of the task difficulty estimation – if the amount of data is sufficient – may be assessed with several models e.g. originating from the Computerized Adaptive Testing (CAT) domain. The Item Response Theory (IRT) provides a range of well-established methods for estimating task difficulty (Rasch, 1960, 1966, 1977) that have been not only utilized in educational (Scheerens et al., 2006), but also medical (Christensen et al., 2013) or marketing applications (Bechtel, 1985). There is however a fundamental limitation of these methods if considering the initial phase of the introduction of the adaptive online learning environment to the public – the requirement of large calibration samples.

There have been several alternative methods for task difficulty estimation examined by the research community in an educational context, e.g. the Elo rating algorithm (Antal, 2013; Klinkenberg et al., 2011; Pankiewicz, 2020a; Pankiewicz & Bator, 2019; Pelánek et al., 2017; Wauters et al., 2012), proportion correct (Antal, 2013; Wauters et al., 2012), or learner feedback method (Chen et al., 2005; Wauters et al., 2012). It has been observed that – for sufficiently big datasets – all of the above-mentioned methods may deliver estimations characterized by reasonably high accuracy. The performance of these methods if the dataset is small has not been extensively researched, especially in the context of environments with the formative assessment approach – where multiple attempts are allowed and feedback is provided after every attempt.

The so called “cold start” problem (Schein et al., 2002) in adaptive learning environments refers to the situation, where little is known about ability of system users and/or difficulty level of items available in the system. Several approaches originating from the machine learning domain (Pliakos et al., 2019; Wei et al., 2017) have been proposed to address this issue in educational context. Recently, also the Elo-based method has been introduced (Park et al., 2019), that integrates the explanatory IRT model. However, if the system is fresh, estimation algorithms are untrained and the introduction of the alternative estimation method is temporary, involvement of sophisticated methods may not be in the focus for the implementing team. Therefore, the question remains if (and to what extent) these less complex methods characterized by low implementation and computational requirements may support the initial phase of the system deployment in adaptive learning environments with the formative assessment approach.
The focus of this research is therefore to examine, to what extent the original Elo rating algorithm may be an appropriate choice for the task difficulty estimation in terms of the so-called “cold start” problem – during the initial phase of the introduction of the adaptive system to the public. The source of the data has been the item-based programming course available on the RunCode online learning environment (Pankiewicz, 2020b). Additionally the Elo estimations have been compared with another simple measure: the Proportion Correct – a method that has been previously found to deliver more accurate estimations than Elo rating algorithm (Antal, 2013; Wauters et al., 2012).

2. Estimating difficulty of an assignment

2.1 Elo rating algorithm

The Elo rating system (Elo, 1978) has been developed for the purpose of measuring strength of players in chess tournaments. The aim of the algorithm is to calculate players’ rating change after every game. That change depends on outcomes of tournament games. Every player is assigned a rating that is usually a number between 1000 and 3000 that is a subject to change after every game. New rating is calculated by a formula:

\[ R_n = R + K(O - P) \] (1)

Where: \( R_n \) is the new value of the rating, \( R \) – the actual rating, \( O \) – game outcome (1 – win, 0 – loss), \( P \) – probability of winning the game and constant \( K \) – the value for chess tournaments is often 32. The probability of winning \( P \) is given as:

\[ P = \frac{1}{1 + 10^{(R_o - R_p) / 400}} \] (2)

Where \( R_o \) is the rating of a player and \( R_p \) is the rating of the opponent. In the context of an online learning environment, we consider a tournament game to be a single submission of a solution, a player – a learner that submits the solution and opponent – a task.

There are three possible outcomes of the chess game (win, lose, draw), but in the context of a learning environment we only consider two outcomes: learner wins if the submission receives the maximum score or learner loses if the submission does not receive maximum score.

2.2 Proportion correct

Proportion correct (PC) – the percentage of correct answers is another simple measure to assess item difficulty calculated for every \( i \)-th item as:

\[ \hat{\beta}_i = 1 - \frac{n_i}{N_i} \] (3)

Where \( n_i \) is the number of correct attempts and \( N_i \) the number of total attempts on the \( i \)-th task.

PC is calculated as the number of learners that solved the task divided by the total number of learners therefore the more learners solved the task in total, the lower the difficulty of that task. According to (Wauters et al., 2012), proportion correct may generate accurate estimations if administered to already 200-250 learners. The accuracy of the method is very high according to several research (Antal, 2013; Wauters et al., 2012).

3. Methodology

3.1 Data

The data originates from the RunCode online learning environment available at https://runcodeapp.com. The online course that supports automated verification of programming tasks has been made available on the RunCode platform as an additional tool in the Introduction to programming course – a mandatory course for the first-semester computer science students at the Faculty of Applied Informatics and Mathematics, Warsaw University of Life Sciences. There have been
50,055 attempts on 76 tasks recorded on the RunCode platform during two editions of the course: 2017/2018 and 2018/2019. There were 299 students in total that used the system during those two editions. Usage of the platform was not mandatory, however majority of students actively participated.

Two methods for estimating task difficulty have been compared in this study: the Elo rating algorithm and Proportion Correct. In order to compare the accuracy of these methods in terms of the cold start problem, reference values of task difficulty have been obtained on the full dataset by the means of the IRT graded response model (Samejima, 1969). The graded response model is suitable for modeling polytomous response data and has been already introduced e.g. for the purpose of knowledge assessment on open-ended tasks with multiple attempts allowed (Attali, 2011).

### 3.2 Sample size computations

The comparison of difficulty estimation methods is presented in regards to the number of attempts. For the $i$-th attempt, calculations include the cumulative number of all trials on tasks that do not exceed $i$ attempts recorded for a single learner on the task. The sample limited by $i$ attempts for a particular learner is selected from all attempts recorded in the data set. No further attempt is contained in the analysis after the attempt that received maximum score. If a learner received a maximum score and submitted another solution, it was not considered in the analysis.

As the outcome of analyses we considered a list of tasks ordered by a difficulty rating. In order to compare difficulty estimations provided by the discussed methods the Pearson’s correlation coefficient has been used.

In order to assess the cold start problem, random samples of sizes $n = 5, 10, 20$ and 50 learners have been drawn from the data set with assurance that at least one attempt has been recorded for every task. The selection procedure has been repeated 1000 times for each sample size. The analysis contains all tasks on which there was at least one attempt, limited to the number of $i$ attempts for every learner. Correlation has been calculated for each sample drawn. Then, the median of obtained correlations has been calculated to acquire the “typical” value. Median has been chosen in order to limit the impact of outliers on final results.

### 4. Results

This analysis is aimed at the evaluation of the Elo rating algorithm in terms of the so-called “cold-start” problem within adaptive online learning environments. Results of the study show that Elo rating algorithm may be a good choice for computing task difficulty estimations during the initial stage of the introduction of the adaptive system to the public, if there is no sufficient amount of data in order to utilize more accurate methods. It has been compared to the Proportion Correct (PC) method – another simple difficulty estimation measure. The Elo method achieves distinctly higher correlation values with the reference data than the PC method for small sample sizes. With increasing size of the sample, the difference between the Elo and Proportion Correct decreases.

The data originates from the programming course available at the RunCode online learning platform used by users of varying programming skills. The course has been made available to the students of the *Introduction to programming* – a mandatory course at the *Faculty of Applied Informatics and Mathematics, Warsaw University of Life Sciences*. Before joining the course students answered in a survey on their self-evaluation of programming skills. More than a half of students declared to have low or very low level of programming knowledge before joining the course.

The average number of unsuccessful attempts on tasks preceding the successful trial was high. Despite the fact that the average difficulty level may be perceived as high, the engagement of platform users was surprisingly high: students did not quickly resign and mostly uploaded another solution.

For the purpose of clarity, results have been presented for the first seven attempts (ca. 75% of all samples). This limitation is reasonable, as the effects visible on the first seven attempts are in general also reflected in the remaining data (e.g. dropout) with the long tail of even more than 50 attempts on a task. The detailed analysis of the submission data has been presented on Table 1.
Table 1. The number of correct and incorrect attempts on assignments. The Total column is the cumulative sum of attempts. The Dropout column is the percentage of students that resigned to take another attempt.

<table>
<thead>
<tr>
<th>Attempts</th>
<th>Incorrect</th>
<th>Correct</th>
<th>Total</th>
<th>Dropout</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8259</td>
<td>5269</td>
<td>13528</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>5623</td>
<td>2389</td>
<td>21540</td>
<td>0.030</td>
</tr>
<tr>
<td>3</td>
<td>4045</td>
<td>1244</td>
<td>26829</td>
<td>0.059</td>
</tr>
<tr>
<td>4</td>
<td>2950</td>
<td>842</td>
<td>30621</td>
<td>0.063</td>
</tr>
<tr>
<td>5</td>
<td>2259</td>
<td>493</td>
<td>33373</td>
<td>0.067</td>
</tr>
<tr>
<td>6</td>
<td>1766</td>
<td>342</td>
<td>35481</td>
<td>0.067</td>
</tr>
<tr>
<td>7</td>
<td>1396</td>
<td>237</td>
<td>37114</td>
<td>0.075</td>
</tr>
</tbody>
</table>

The sampling procedure presented in the section 3.2 has been performed on the following sizes of the sample: n = 5, 10, 20 and 50. Results of the analysis – Pearson’s correlation values between reference values and the Elo and PC difficulty estimations for the first seven attempts – have been presented on the Figure 1.

Figure 1. Correlation of the item difficulty estimates with the reference values for the range of attempts between 1 and 7 and sample size of 5, 10, 20 and 50. Estimation methods: Elo rating algorithm, proportion correct (PC).

The highest value of the correlation for the size of the sample n = 5 between the Elo estimation and the reference values: cor_{ELO5S} = 0.702, PC: cor_{PC5S} = 0.644. For the size of the sample n = 10, the highest observed correlation value cor_{ELO10S} = 0.784, PC: cor_{PC10S} = 0.726. For n = 20, the highest value cor_{ELO20S} = 0.852, PC: cor_{PC20S} = 0.821. For n = 50, the highest value cor_{ELO50S} = 0.905, PC: cor_{PC50S} = 0.896. For all of analyzed sizes of the sample, the Elo rating algorithm achieved higher values of correlation than the proportion correct method (Table 2).
Table 2. Correlation of the item difficulty estimates with the reference values for the range of attempts between 1 and 7 and sample size of 5, 10, 20 and 50. Estimation methods: Elo rating algorithm, proportion correct (PC).

<table>
<thead>
<tr>
<th>attempts</th>
<th>Size = 5</th>
<th></th>
<th>Size = 10</th>
<th></th>
<th>Size = 20</th>
<th></th>
<th>Size = 50</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Elo</td>
<td>PC</td>
<td>Elo</td>
<td>PC</td>
<td>Elo</td>
<td>PC</td>
<td>Elo</td>
<td>PC</td>
</tr>
<tr>
<td>1</td>
<td>0.378</td>
<td>0.343</td>
<td>0.433</td>
<td>0.395</td>
<td>0.484</td>
<td>0.439</td>
<td>0.533</td>
<td>0.487</td>
</tr>
<tr>
<td>2</td>
<td>0.649</td>
<td>0.616</td>
<td>0.728</td>
<td>0.697</td>
<td>0.797</td>
<td>0.780</td>
<td>0.859</td>
<td>0.847</td>
</tr>
<tr>
<td>3</td>
<td>0.696</td>
<td>0.644</td>
<td>0.779</td>
<td>0.726</td>
<td>0.849</td>
<td>0.819</td>
<td>0.905</td>
<td>0.892</td>
</tr>
<tr>
<td>4</td>
<td>0.702</td>
<td>0.637</td>
<td>0.784</td>
<td>0.724</td>
<td>0.852</td>
<td>0.821</td>
<td>0.902</td>
<td>0.896</td>
</tr>
<tr>
<td>5</td>
<td>0.697</td>
<td>0.627</td>
<td>0.777</td>
<td>0.711</td>
<td>0.845</td>
<td>0.809</td>
<td>0.894</td>
<td>0.885</td>
</tr>
<tr>
<td>6</td>
<td>0.693</td>
<td>0.620</td>
<td>0.773</td>
<td>0.705</td>
<td>0.838</td>
<td>0.800</td>
<td>0.885</td>
<td>0.878</td>
</tr>
<tr>
<td>7</td>
<td>0.678</td>
<td>0.607</td>
<td>0.756</td>
<td>0.688</td>
<td>0.821</td>
<td>0.783</td>
<td>0.867</td>
<td>0.863</td>
</tr>
</tbody>
</table>

5. Summary and discussion

The aim of this study was to present the Elo rating algorithm as a tool for assessing the task difficulty in terms of the so-called “cold-start” problem. This analysis has been performed on the real data originating from the online item-based Introduction to programming course available on the RunCode platform: the online learning environment where multiple attempts were allowed and feedback was provided after every attempt. The so-called “cold start” problem refers to the initial stage of the introduction of an adaptive learning system to the public, where little is known about system users and/or available items. Until a sufficient amount of data is gathered and item bank calibration may be performed, usage of methods e.g. originating from the area of the Computerized Adaptive Testing domain becomes a hurdle. Therefore alternative, simpler methods of difficulty estimations, such as Elo rating algorithm may be considered as a temporary (or sometimes permanent) solution. Results of this study showed that the Elo rating algorithm achieved the correlation of 0.702 with the reference difficulty estimation values obtained by the means of the IRT graded response model for the size of the sample n = 5. For the size of the sample n = 10: cor. = 0.784, for n = 20: cor. = 0.852, and for n = 50: cor. = 0.905. Estimations obtained by the Elo method outperform values calculated by the proportion correct measure for sizes of the sample n = 5, 10 and 20. As the size of the sample increases, difference between estimations calculated by the Elo rating algorithm and Proportion Correct method decreases.

The first conclusion drawn from results of this study is that the introduction of the Elo rating algorithm for the purpose of assessing the task difficulty at the initial stage of the introduction of the adaptive system to the public may be a reasonable choice. Already for the size of the sample as small as n = 5, the method achieves reasonably high correlation of 0.702. For larger size of the sample n = 50, the Elo rating algorithm achieves the correlation value of 0.905, however Proportion Correct achieves comparable value: 0.896. Both methods are characterized by low computational requirements, and – compared to e.g. expert rating or learner feedback – do not require additional human engagement. There is also an additional benefit: the complexity of the method is low and therefore its implementation may be easily carried out.

The second conclusion is that the Elo rating algorithm performs better than the Proportion Correct method for smaller sizes of the sample. The Elo method quickly “learns” from user submission results and therefore much quicker adjusts its difficulty rating. As the number of recorded submissions increases, this benefit of the Elo method is less visible in the outcome and estimations of both PC and Elo methods become similar. There is no direct comparison to be established with previous research, as it has been mainly focused on the summative assessment: in the study of (Wauters et al., 2012) both the Proportion Correct and Elo method similarly estimate reference values for small size of the sample with small advantage of the PC method, and in the study of (Antal, 2013) the Proportion Correct method delivers more accurate estimations.

Main limitations of this study refer to the aspect of high engagement observed within analyzed group of platform users. The observed dropout rate has been very low and may be a result of implemented gamification elements, but also may result from the fact that system users were university’s computer science students. Although the usage of the platform was not mandatory, and
results obtained within the platform did not impact final grade, the motivation of students may have been higher than of an average group of people interested in gaining experience in programming. It is more probable that these observations will be replicable in university settings, than on any publicly available online learning platform.

References


Design of a Self-Reflection Model in GOAL to Support Students' Reflection

Huiyong LIa*, Rwitajit MAJUMDARB, Yuanyuan YANGa & Hiroaki OGATAb

aGraduate School of Informatics, Kyoto University, Japan
bAcademic Center for Computing and Media Studies, Kyoto University, Japan

* lihuiyong123@gmail.com

Abstract: GOAL (Goal Oriented Active Learner) system is a platform to support the development of students' self-direction skills using a data-driven feedback loop. It integrates students’ learning activity data from the learning management system (LMS) and physical activities data from the mobile health applications or wearable devices. These activity data provide a data-rich context for students to train their self-direction skills, such as planning, monitoring and reflection. The reflection skill plays an important role in the self-directed cycle, and therefore students should be given more opportunities to practically execute the skill and further improve it. This paper proposes a self-reflection model to support the acquisition of reflection skills in students’ self-directed learning and physical activities. The model is built on reflective process and strategies in self-direction and self-regulation. It allows students to reflect on their plans and achievements in the planning-monitoring-reflection process and enables them to consider their reflective strategies. The conceptual contribution of this paper is applying reflection theory to the practice in self-directed activities through computer based scaffolds.

Keywords: Self-reflection, student reflection, learning analytics, self-direction, GOAL

1. Introduction

Self-direction Skills (SDS) are acquired through experience, training, and effort. The gain of experience and training depends on the degree to which learners engage in volitionally initiated processes. Since it is a cognitively and behaviorally complex task to execute self-direction, more opportunities that learners engage in self-direction may benefit the acquisition of SDS.

Although there are multiple approaches to capture data on learners’ self-direction or self-regulation, self-report measures have still stayed dominant so far. The recent availability of large and fine-grained datasets has led to investigate self-regulation by applying learning analytics (LA). The assessment of frequencies and sequences of regulatory activities in learning environments provides a novel perspective on self-regulation that complements and potentially supersedes traditional self-report measures (Bannert & Sonnenberg, 2014; Li et al., 2018). On the other hand, the increased availability of activity tracking data gives individuals more opportunities for establishing benchmarks in objective metrics and improving achievements through the experience of reality (Swan, 2013). The research and design of data quantification have grown as an interest area in information and learning sciences (Lee, 2019).

This leads us to build the GOAL system to support the development of students’ SDS using a data-driven feedback loop (Majumdar et al., 2018). The GOAL system not only leverages the rapidly increasing activity data but also provides computer based scaffolding to foster students’ SDS in the self-direction cycle. The reflection is a key phase in the self-direction cycle. In this paper, we introduce how to build a self-reflection model and support the acquisition of reflection skills.

2. Related Work

2.1 Self-Direction Skills

According to P21 (Partnership for 21st Century Skills, 2016) framework, Initiative and Self-Direction
requires monitoring one's understanding and learning needs, demonstrating initiative to advance professional skill levels, defining, prioritizing and completing tasks without direct oversight and demonstrating commitment to lifelong learning. It requires learners to handle multiple environments, goals, and tasks while understanding and adhering to organizational or technological constraints of time, resources, and systems. The framework gives a general criterion for a self-directed learner.

Self-directed learning (SDL) and self-regulated learning (SRL) are two terms most frequently used in today’s educational discourse on learning process (Brockett & Hiemstra, 2018; Candy, 1991; Winne et al., 2006; Zimmerman, 2008). Literature highlights their commonality and differences (Saks & Leijen, 2014). Both SDL and SRL have 4 key phases: Task definition – Setting goals and Planning – Enacting strategies – Monitoring and Reflecting.

Technological innovation in the field of data logging and rapidly increasing digital world have expanded the intersection of SDL and SRL, so that, the processes of executing and developing SDL and SRL can be captured. We have proposed DAPER (Data collection - Analysis - Planning - Execution monitoring - Reflection) model which synthesizes the SDL and SRL models to conceptualize data-driven self-direction skill execution and acquisition (Majumdar et al., 2018). It is a process model with five phases, the initial phase of data collection which gives learners the initiative in their contexts, followed by the other four phases: data analysis, planning, execution monitoring and reflection. Figure 1 shows the DAPER phases with example from the context of learning and physical activities.

Figure 1. DAPER Model of Self-Direction Skills Execution and Acquisition (Majumdar et al., 2018)

2.2 Self-Reflection in Self-Direction and Self-Regulation

There are many indications that reflection plays an important role in learning processes (Boud et al., 1985; Holmberg, 2005; Hammond & Collins, 2013), and that reflection also constitutes an important factor in the acquisition of self-direction and self-regulation (Kuiper & Pesut, 2004; Nesbit, 2012). Reflection can be conceived of as a skill or strategy that operates on other strategies, such as planning, monitoring. Learners should reflect on not only what they had learned (monitoring/reflection) but also how they learned (planning/monitoring). By reflecting on one's own learners become aware of their learning processes and possible alternative strategies. During self-directed activities, reflection becomes of greater importance for learner success when the process is less externally guided (e.g., by the teacher). Learners must then manage their learning to a greater extent, making reflection more critical. Therefore, learners require more strong support for self-reflection.

2.3 Support Self-Reflection in Self-Direction and Self-Regulation

While there is a broad consensus that reflection is a crucial factor for the improvement of students' self-direction and self-regulation, it is also found that the majority of students do not reflect deeply on their learning processes in educational practice (Zimmerman, 2006). In order to engage students in effective self-reflection, a scaffolding environment for training can be considered. Regarding this support environment, the question is: what should provide to self-reflection for students? Students may
lose their directions if without reliable, revealing and relevant data that support effective reflection. Following the learning analytics process model learners need to translate awareness into action (Bodily et al., 2018). They need a representative reference frame to interpret the data (Wise, 2014). Both the contextual activity data and trace data of self-direction behavior can be valuable ways to create such a reference frame.

Furthermore, reflection journal writing is effective in promoting self-reflection and learning (Lew & Schmidt, 2011). Students can reflect on their process and achievement through a structured reflection journal. Prompting has been identified as a promising method to evoke these reflective actions. Therefore, in this study a computer based scaffold is integrated using a reference frame, a reflection journal, and reflection prompts. It is designed to support and stimulate students to reflect on their learning processes and achievements.

3. Design of the GOAL System

The design and implementation of the GOAL system is shown in Figure 2. The GOAL system integrates data during learners’ learning and physical activities, tracks the interactions between learners and system, and implements the DAPER model with the functionalities required in each phase. Learners can link automatically their learning activity data from the LMS and other linked e-learning tools, such as digitized reading logs, answers of quizzes, and status of course assignments (Flanagan & Ogata, 2018). Learners can also synchronize physical activity data directly from mobile health apps or platforms for wearable devices, such as data from runs, workouts, sleep, steps taken, weight, heart rate, and calories burned. Furthermore, the interactions between learners and the GOAL system are logged as eXperience API (xAPI) statements in the GOAL server. This system grounds the theory of SDS and enables learners to develop the skills in the context of learning and physical activities.

![Figure 2. The Design and Implementation of the GOAL System](image)

In the pre-planning phase, students could collect and analyze their own activity data from the learning or health platform. Once they have set a plan for a learning or physical activity, they start to monitor the progress and reflect on the learning processes and outcomes in the post-planning phase (Li et al., 2019). The students have active involvement in these phases since they increase their ownership by interacting with their own activity data and plans. They can create learning or physical activity plans, such as an extensive reading activity in English in weekly scale or a running activity in daily scale.

4. Design of the Self-Reflection Model

Our proposed self-reflection model is shown in Figure 3. First, a reference panel is shown, which consists of the detail of plan, target achievements, and activity achievements. Second, a structured reflection journal is provided to let students rate the degree of plan difficulty, the target achievement rate, and the effort to achieve the plan. An unstructured comment is also added in the reflection journal. Third, a skill diagnosis of self-reflection skills is performed using a 5-point scoring rubric. Finally, an adaptive feedback is generated based on the diagnosed score. The model is executed by students in self-reflection process in an iterative cycle.
4.1 Reference Panel

The reference panel contains the detail of plan, target achievements, and activity achievements. The detail of plan has these items: plan name, activity type, start date, end date, frequency, target value, target unit, and notes. The target achievement shows the days when the learner achieved the target value and the completion rate within the planned duration. The activity achievement shows the average and total amount of the planned activity.

4.2 Reflection Journal

The reflection journal contains the degree of plan difficulty, the target achievement rate, and the effort to achieve this plan. These indicators can be rated as a 3-point score. Also, an unstructured thought could be input in an additional comment area by learners, which can be the current problems, specific strategies, or further actions.

4.3 Skill Diagnosis using 5-point Scoring Rubric

The self-reflection skills are measured by a 5-point scoring rubric. Table 1 indicates the scoring rubric for self-reflection skills with the score, criteria and interaction objects. Five levels of self-reflection skills are diagnosed: never reflect, reflect on personal plan only, reflect on personal plan and achievement, reflect by self-rating but no comments, and reflect by self-rating and further comments.

<table>
<thead>
<tr>
<th>Score</th>
<th>Criteria</th>
<th>Interaction Objects</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Reflect by self-rating and further comments</td>
<td>Reflection journal</td>
</tr>
<tr>
<td>3</td>
<td>Reflect by self-rating but no comments</td>
<td>Reflection journal</td>
</tr>
<tr>
<td>2</td>
<td>Reflect on personal plan and achievement</td>
<td>Reference panel</td>
</tr>
<tr>
<td>1</td>
<td>Reflect on personal plan only</td>
<td>Reference panel</td>
</tr>
<tr>
<td>0</td>
<td>Never reflect</td>
<td></td>
</tr>
</tbody>
</table>

4.4 Adaptive Feedback based on Skill Scores

Learners are classified into 5 groups based on the diagnosed skill scores. They are given adaptive feedback through feedback prompts (see Table 2). The feedback prompts are actionable suggestions which support learners continuously to improve their reflection skills.
Table 2: Adaptive Feedback for Learners based on Skill Scores

<table>
<thead>
<tr>
<th>Score</th>
<th>Self-Reflection Skill Level</th>
<th>Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Reflect by self-rating and further comments</td>
<td>Well done! You got a great reflection skill</td>
</tr>
<tr>
<td>3</td>
<td>Reflect by self-rating but no comments</td>
<td>Great! Then try to reflect on your strategies and record it into comments</td>
</tr>
<tr>
<td>2</td>
<td>Reflect on personal plan and achievement</td>
<td>Please try to rate by yourself about your plan and achievement</td>
</tr>
<tr>
<td>1</td>
<td>Reflect on personal plan only</td>
<td>Please check your achievement</td>
</tr>
<tr>
<td>0</td>
<td>Never reflect</td>
<td>Please check your plan</td>
</tr>
</tbody>
</table>

5. Self-Reflection Model-Based Interface in GOAL System

The self-reflection model-based interface in GOAL system is shown in Figure 4. Three components of self-reflection model are provided: reference panel (plan detail and achievement), reflection journal, and adaptive feedback. The contextual activity in this interface is extensive reading in English. The learner has created a daily reading plan for extensive reading, and could review the achievements since the plan was finished.

![Figure 4. Self-Reflection Model-Based Interface in GOAL System](image)

6. Conclusion and Discussion

In this paper we proposed a novel model for self-reflection and support the acquisition of reflection skills in the context of self-directed activities, like e-book reading, walking, or running. The reference panel of this model triggers learners to reflect what they had learned (monitoring/reflection) and how they learned (planning/monitoring). The structured reflection journal of this model guides learners to rate key indicators in the planning-monitoring-reflection process and lets them reflect deeply and critically. The skill diagnosis and adaptive feedback help learners to further understand their learning processes and outcomes to improve their self-reflection skills. The benefits of using the model are to facilitate the transfer of skills between the system and learners. The conceptual contribution of this paper is applying reflection theory to the practice in self-directed activities through computer based scaffolds.

The proposed model is executed as an iterative cycle to facilitate a continuous improvement in self-reflection. Furthermore, it involves other phases of self-direction process, such as planning and monitoring. For instance, it can enhance subsequent planning and monitoring phases and then further reflection. Therefore, the model can influence the whole cycle of self-direction process and foster learners’ SDS. The effect of the proposed model will be examined in the future.
Acknowledgements

This research was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (S) Grant Number 16H06304, NEDO Special Innovation Program on AI and Big Data 18102059-0, JSPS KAKENHI Grant-in-Aid for Early-Career Scientists 20K20131.

References


Automatic Feedback Models to Students Freely Written Comments

Jihed MAKHLOUF & Tsunernori MINE

Kyushu University, Fukuoka, Japan

jihed.makhlouf@m.ait.kyushu-u.ac.jp, mine@ait.kyushu-u.ac.jp

Abstract: Teachers and professors, in different educational institutions, always wanted to grasp the learning experience of their students so they can give them proper guidance and intervene when it is necessary. However, tracking every student can be very challenging. In this context, we created an online questionnaire and asked students to fill it after each lesson. The Professor read what the students had written to get a better understanding of their learning experience and to reply to them with the adequate guidance. However, it turns out that professors quickly become overwhelmed and students have to wait longer before receiving any feedback. In this paper, we describe our method of building an automatic feedback model that will be used to help professors. We tried two approaches of building the models with or without a padding of the context. Empirical results show that our padded models can achieve 0.664 micro F-score.

Keywords: Educational Data Mining, Natural Language Processing, Genetic Programming, Comments Mining

1. Introduction

Providing students with a better learning experience have always been an important topic in learning science and educational technology. Thanks to the continuous advances in educational technology, more educational institutions are adopting educational software systems. Indeed, the usage of such systems opens-up countless opportunities of gathering and analyzing insightful data. Also, it allows building different sophisticated models that help improving the students learning experience (Dietz & Hurn, 2013). Furthermore, predicting students’ performance is a topic of interest to many researchers. In fact, many different means are used to assess the students’ performance. Some of them rely on careful observations during class time, while others are more explicitly elaborated such as test scores and questionnaires (Minami & Ohura, 2015). Using the traditional exercise assessment, test scores and attendance is very handy and helpful, but sometimes they are not enough to fully grasp the whole range of students’ behavior and learning experience (Yamtim & Wongwanich, 2014). Therefore, it is important to find different ways of gathering students’ data that allow the production of high-level research and prediction models for students’ performance, behaviors and affects. These predictive models rely on different types and forms of stored data. Indeed, some sources of very valuable data are questionnaires and surveys. They have been used for a long time, however, research using solely data coming from questionnaires is still limited compared to other sources of data. For instance, (Bachtiar, Kamei, & Cooper, 2011) designed a questionnaire that measures the students affect such as personality, motivation and attitude, then they built a predictive model of students’ english language aptitude based on reading, speaking and writing independently. In a different context, (Jiang, Syed, & Golab, 2016) used a large collection of course evaluation survey for undergraduate and built a predictive model using linear regression to extract the aspects that influence the evaluation of the course. Predictive models using data gathered from questionnaires are not abundant. It is even more rare to find research topics that use solely textual data coming from questionnaires. For example, (Sliusarenko, Clemmensen, & Erbsboll, 2013) used the textual data gathered from a course evaluation rating survey. The survey’s textual data consists of open-ended comments. Then the authors extracted the most important aspects in the students’ comments and investigated how they influence their rating of the course. (Minami & Ohura, 2013) used the term-end questionnaire to extract students’ textual input. They combined the textual data with other sources of data like attendance, test scores and
homework evaluation scores and identified the common writing characteristics of highly successful students.

In a different context, (Goda & Mine, 2011) designed a questionnaire where students are requested to self-reflect on their learning experience using freely written comments. The survey is conducted after each lesson. The authors also proposed the PCN method. PCN is the abbreviation of Previous, Current and Next. It provides the ability to acquire temporal information of each student’s learning activity related to the corresponding lesson. The first subset P (Previous) covers all the student’s activities prior to the lesson. It can be in the form of preparation of the actual lesson or a review of the previous lesson. The second subset C (Current) relates to all activities made during the class. It particularly covers the student’s understanding of the content of the lesson, the problems that he/she has faced and the activities that involve teamwork or communication with peer classmates. Finally, the subset N (Next) encapsulates the students’ comments about plans to review the actual lesson and prepare for the next lesson. The authors declared that the PCN method incited students to improve their self-reflection on their learning environment and to better strategize on their learning activities’ planning. Meanwhile, teachers gather valuable data about their students learning experience.

Several subsequent researches were made using the PCN method, mainly to predict students’ performance and grades (Sorour, Goda, & Mine, 2017 ; Sorour, Goda, & Mine, 2015 ; Sorour, Mine, Goda, & Hirokawa, 2014). While the prediction models using the students’ freely written comments based on the PCN method achieved robust performances, the professors still had to read a large amount of comments to grasp the students’ learning experience and to provide their feedback to each student after each lesson. And in the other side, the students had to check and wait for the professor’s reply. Therefore, this method is not optimal. In our previous work (Makhlof & Mine, 2020), we automated the process of assessing the students’ comments. Therefore, professors can get a quick idea about the students’ learning experience using prediction models and without the need to read carefully all the students’ comments. Therefore, the aim of this research is to address the feedback issue of this method. In fact, students have to wait for some time to get a reply from their professor. Meanwhile, the professor has to read carefully each comment in order to provide a proper reply. Hence, our objective of building an automatic feedback model that will be used to replace the professor and give feedback in real-time to students.

2. Methodology

2.1 Data acquisition

We gathered the data using a PCN-based questionnaire operated during the 2017 and 2018 Functional Programming course for undergraduate students. Totally, we count 109 different students that were enrolled in these courses. In the dataset files, each row consists of a student’s “response” to the questionnaire after a lesson. Each one of these “responses” is composed by 5 freely-written comments related to the 5 predefined questions following the PCN method. More details about the 5 predefined questions are exposed in the Table 1.

<table>
<thead>
<tr>
<th>Subset</th>
<th>Question</th>
<th>Example of comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>P (Previous)</td>
<td>What did you do to prepare for this lecture?</td>
<td>I read the syllabus.</td>
</tr>
<tr>
<td>C (Current)</td>
<td>Do you have anything you did not understand? Any questions?</td>
<td>I had problems when installing the environment.</td>
</tr>
<tr>
<td></td>
<td>What are your findings in this lesson?</td>
<td>I understood the basics of functional programming.</td>
</tr>
<tr>
<td></td>
<td>Did you discuss or cooperate with your friends?</td>
<td>I talked with my friends about errors in my computer.</td>
</tr>
<tr>
<td>N (Next)</td>
<td>What is your plan to do for the next lecture?</td>
<td>I will do my best to avoid my errors and submit the report.</td>
</tr>
</tbody>
</table>
In fact, as shown in Table 1, in the subset P, students describe their learning activities prior to the lesson. In the subset C, we have 3 different questions. Firstly, students describe their problems and which content they did not understand well. The second question is about their discoveries during the lesson and finally they report their interactions and cooperation with their peer classmates. In the subset N, students detail their plans for the next lesson. Each comment related to each question has its own feedback. Therefore, when dealing with the dataset we split the “responses” into individual comments. And after a first cleanup of invalid or empty comments we ended up having an overall of 2547 comments regardless of the question.

2.2 Manual feedback labeling

Since our objective is to build an automatic feedback model, we have to generate a dataset first. For this, we asked two students in their master program to provide a well thought feedback to each comment. Since the comments are gathered from an undergraduate course, the task was not hard. Moreover, a deep understanding of the course was not needed to accomplish this task, since the comments are more related to the students’ self-assessment of their learning activities than to the in-depth knowledge of the content of the course.

2.3 Initial data exploration

After some cleaning and preprocessing we end up with 112 unique feedbacks. Therefore, it is considered as a multi class classification problem. To deal with these types of problems it is necessary to verify the class balance. In fact, the feedback classes are unbalanced, as more than a fifth of the classes are redundant. However, it is not a steep unbalanced distribution, since building a model that constantly predicts the most frequent class will be accurate only 11% of the time. Therefore, the most frequent feedbacks are not extremely dominant. Moreover, when we check the students’ comments as well, we find that out of the 2547 comments there is 1731 unique comments. It means that 67% of the comments were not duplicated.

2.4 Text preprocessing and features engineering

The textual data are written in Japanese. Therefore, the preprocessing steps have to be done accordingly. Moreover, since it is a programming course, students frequently used punctuation signs and special characters in their comments. Also, English words were used in between Japanese text. So, the preprocessing phase consisted mainly of removing line breaks, redundant or extra spaces and tabs. English words were normalized. For the Japanese text, the first step was to normalize it to avoid problems such as half-width and full-width characters’ troubles. When the text is cleaned and normalized, we use MeCab for the parsing and POS (Part of Speech) tagging. MeCab\(^1\) is a dictionary-based Part-of-Speech and Morphological Analyzer of the Japanese language. After that, to process the textual data, it has to be transformed into numerical values. There exist different encoding techniques, we picked two widely adopted methods which are TF-IDF and Doc2Vec. When we used TF-IDF, we generated the unigrams, and bi-grams of the textual data which created vectors of 3326 dimensions. For the Doc2Vec, we generated vectors of 300 dimensions.

3. Experiments

3.1 Models building approaches

In this work we investigate two different approaches for building our automatic feedback models. In our baseline approach we use the student comments without incorporating the context of the question. In our second approach we try to incorporate information about the question as part of the comment. We used a simple padding of the type of the question before each comment. For example, if the student commented: “I reviewed the content and practiced at home” when answering the question “What did

---

\(^1\) https://taku910.github.io/mecab/
you do to prepare for this lecture?”, then we transform the comment by adding a padding like: “<preparation> I reviewed the content and practiced at home”. It is worth mentioning that the padding phase is done before the tokenization phase using MeCab, therefore the padding will also be part of the preprocessed text. We also investigate which textual encoding technique gives us better results. In the end, we end up comparing 4 different models, as summarized in Table 2.

Table 2. Summary of the different models

<table>
<thead>
<tr>
<th>Model name</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>TF_Simple</td>
<td>Baseline approach using the TF-IDF text encoding technique</td>
</tr>
<tr>
<td>D2V_Simple</td>
<td>Baseline approach using the Doc2Vec weights</td>
</tr>
<tr>
<td>TF_Padded</td>
<td>Padding the comment with the question type and encode the text using TF-IDF</td>
</tr>
<tr>
<td>D2V_Padded</td>
<td>Padding the comment with the question type and encode the text using Doc2Vec</td>
</tr>
</tbody>
</table>

3.2 Hyper-parameters tuning using Genetic Programming

Briefly, genetic programming is a technique derived from genetic algorithms in which instructions are encoded into a population of genes. The goal is to evolve this population using genetic algorithm operators to constantly update the population until a predefined condition is met (Koza., 1992). When using genetic programming for machine learning optimization, each individual of the population represents a pipeline that holds a machine learning technique with its hyper-parameters. We use the pipeline score, like the accuracy, as the objective function that has to be maximized. This step was done using the TPOT python library (Olson, Bartley, Urbanowicz, & Moore, 2016). By the end of this process, we will find the machine learning pipeline that has the best prediction score. Therefore, our objective from the optimization step is to find for each approach its best machine learning pipeline, then ultimately comparing them to each other.

3.3 Evaluation metrics

To determine which model gives the best results we will use a mix of evaluation metrics. Since we are facing a multiclass classification problem, we need to use class-wise metrics. Therefore, we used the macro F-score and micro F-score. However, the nature of our textual classes makes it incomplete to use only these metrics. In fact, many feedback classes are a minority and have only one occurrence. But, many of them are semantically similar even if they belong to different classes. Therefore, we add another metric that we measure to quantify how similar are the predicted feedback and the true feedback. This measure will allow us to tell how well the model can give a similar feedback if it does not predict the correct one. We used the Word Mover’s Distance (WMD) which is a measure of semantic similarity between two sentences. It leverages the power of the word embedding such as Word2Vec to measure a distance value. The smaller the value the better it is. And a value of zero means that the two sentences are identical. We measure the average WMD and the max WMD for each model.

3.4 Experimental results

We firstly proceed to the optimization phase where we use genetic programming to find the best machine learning pipeline. The results of the optimization phase are shown in Table 3.

Table 3. Results of the optimization process

<table>
<thead>
<tr>
<th>Model’s name</th>
<th>Best method</th>
<th>Best Pipeline Accuracy Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>TF_Simple</td>
<td>Support Vector Machine</td>
<td>0.433</td>
</tr>
<tr>
<td>D2V_Simple</td>
<td>Random Forest Classifier</td>
<td>0.379</td>
</tr>
<tr>
<td>TF_Padded</td>
<td>XGBoost</td>
<td>0.650</td>
</tr>
<tr>
<td>D2V_Padded</td>
<td>Random Forest Classifier</td>
<td>0.519</td>
</tr>
</tbody>
</table>
As shown in Table 4, in the baseline approach, the optimization process found that the Support Vector Machine had the best accuracy result in the TF-IDF based model with a score of 0.433. When using the Doc2Vec, Random Forest Classifier was the best pipeline by reaching 0.379 accuracy score. However, when we integrate the context of the question in the comment we have an improvement in the accuracy score for both models. In fact, in the TF-IDF model we found that the XGBoost algorithm achieved the best results with an accuracy score of 0.650 and when using the Doc2Vec, we have again the Random Forest Classifier as the best pipeline, attaining 0.519 accuracy score.

Since we now have the best machine learning techniques with the best hyper parameters, we proceed now to train and validate them accordingly. We originally split the dataset into 80% training and 20% testing.

The results of the validation phase are shown in the Table 4. We can see that the model using TF-IDF vectors and padding the question type in the comments achieved the best scores across all validation measures. In fact, it has reached 0.247 in the Macro F-score, 0.664 in the Micro F-score, 0.107 in the average word mover’s distance and a maximum WMD of 0.507. On the other hand, the baseline approach using Doc2Vec had the worst performances by having the lowest Macro F-score of 0.106, the lowest Micro F-score as well, going down to 0.356. Its average WMD is the highest attaining 0.212 which means that when it does not classify the feedback correctly, it still does not choose a close enough class. The worst value of Max WMD is achieved by the baseline model using TF-IDF.

Table 4. Validation scores for all models

<table>
<thead>
<tr>
<th>Model’s name</th>
<th>Macro F-score</th>
<th>Micro F-score</th>
<th>Avg WMD</th>
<th>Max WMD</th>
</tr>
</thead>
<tbody>
<tr>
<td>TF_Simple</td>
<td>0.133</td>
<td>0.429</td>
<td>0.184</td>
<td>0.543</td>
</tr>
<tr>
<td>D2V_Simple</td>
<td>0.106</td>
<td>0.356</td>
<td>0.212</td>
<td>0.529</td>
</tr>
<tr>
<td>TF_Padded</td>
<td><strong>0.247</strong></td>
<td><strong>0.664</strong></td>
<td><strong>0.107</strong></td>
<td><strong>0.507</strong></td>
</tr>
<tr>
<td>D2V_Padded</td>
<td>0.158</td>
<td>0.467</td>
<td>0.136</td>
<td>0.514</td>
</tr>
</tbody>
</table>

4. Discussion

The experimental results show that adding the context of the question in the comment resulted in a significant boost in the models performances. In fact, the best results were achieved by the padded model using TF-IDF, and the second best is the padded model using Doc2Vec. We also notice that Doc2Vec was outperformed by TF-IDF in most cases, both in the baseline approach and the padding approach. In a first sight, the F-score measures do not seem high, but they are good considering that we are dealing with a multiclass classification problem having 112 different classes. Moreover, we had to deal with the problem of imbalanced data even if no class is particularly dominant. Hence, the addition of the WMD as another metric to quantify the ability of the model to choose a close enough class if it does not correctly predict the class. Also, we found that a simple padding had a great effect on improving the performance. Incorporating the type of the question in the comment was proven to be effective and that shows the importance of the context of the comment in predicting the correct feedback.

Moreover, during the process of finding the best machine learning method, the usage of genetic programming was effective since it can be considered as a heuristic-based grid search, where we do not evaluate all possibilities but rather keep only the promising ones according to a certain strategy.

5. Conclusion

In this paper, we explained the steps we had taken to build an automatic feedback model given students’ comments. Following the PCN method, students have to provide comments to 5 predefined questions. We gathered the adequate feedback and transformed the problem into a multiclass classification. We investigated two different techniques to encode the text into numerical values, the TF-IDF and Doc2Vec. We also investigated the effect of adding the context of the question in the comment itself. The addition of the context of the question as a padding improved significantly the performance of the models. While it is very helpful for the actual situation, it is not very scalable since if we add a different question in the questionnaire, the model has to be redone to include the newly added question.
In future improvements, we will gather more students’ comments from more recent classes. It will help having more robust models. Moreover, we plan to use different techniques that are also useful in this situation such as clustering. Also, thanks to the advances in NLP, we are able to use more recent and better language models to achieve our goals.

Acknowledgements

This work was partially supported by JSPS KAKENHI Grant Numbers JP18K18656, JP19KK0257, JP20H04300, and JP20H01728.

References


Exploring Temporal Study Patterns in eBook-based Learning

Gökhan AKÇAPINAR*, Mohammad Nehal HASNINE, Rwitajit MAJUMDAR, Mei-Rong Alice CHEN, Brendan FLANAGAN & Hiroaki OGATA

* Department of Computer Education & Instructional Technology, Hacettepe University, Turkey
# Research Center for Computing and Multimedia Studies, Hosei University, Japan
# Academic Center for Computing and Media Studies, Kyoto University, Japan
*gokhana@hacettepe.edu.tr

Abstract: In this study, approximately 2 million click-stream data of 1346 students in the eBook platform were analyzed aiming to explore the temporal study patterns of the students followed during the lectures. The data used in the study collected from Kyushu University, Japan with the help of a digital textbook reader called BookRoll. Students used BookRoll for reading learning materials in and out of the class. To analyze the data we first, converted reading sessions into the sequence data which represents student’s weekly reading behavior, then we clustered students based on their study patterns. Our results revealed that three groups of students can be extracted with similar study patterns. Most of the students in Cluster 1 viewed the learning materials only during the class, without previewing and reviewing them. Students in Cluster 2 previewed the learning materials before the class, viewed learning materials during the class, and also reviewed after the class. Students in Cluster 3 viewed the learning materials during the class in the beginning but they became inactive over the period of time (week by week). Our study also showed how learning analytics can be used to understand students’ study patterns which are difficult to do with self-report data. These results can help instructors while designing their courses.

Keywords: sequence mining, clustering, reading logs, eBook, educational data mining, learning analytics

1. Introduction

Pre-class reading and after class repetition are crucial for students in higher education to understand the subject being taught in class and transfer their knowledge to other domains. Typically, the purpose of the pre-class reading assignment is to expose students to background knowledge that will be useful in an upcoming class discussion or to introduce a topic that will be presented more directly by the instructor (Tomasek, 2009). Completing this pre-class work helps students to be more engaged in the in-class learning process (Ripley, 2007). In another study, researchers found that students’ exam performance significantly improves by nearly 12% in the flipped-format course, due in part to students interacting with course material in a more timely and accurate manner (Gross, Pietri, Anderson, Moyano-Camihort, & Graham, 2015). Although pre-class reading has many advantages on students’ learning, studies have shown that students frequently do not read their textbooks before the class (Ruscio, 2001), moreover, most of the students are not reading the textbook at all (Lieu, Wong, Asefirad, & Shaffer, 2017).

The advancement of online learning technologies such as ITSs, digital eBook systems, MOOC’s, etc. opened the doors to the learners to gain new knowledge. Regardless of learners' knowledge, motivation, or engagement level, learners get the flexibility to engage with the learning systems by navigating through various learning materials (Boroujeni & Dillenbourg, 2018). As a result, learners leave various study patterns which may tell a lot about learners' learning processes. Analyzing study patterns is important and has gained significant attention of educational researchers because hidden in study patterns can provide many important insights about learners and learning environments.

In the present study, we focus on students’ temporal study patterns by analyzing a large amount of click-stream data collected from university students related to their pre-class, in-class, and after-class
reading behaviors. As mention by Knight, Friend Wise, and Chen (2017) learning is a process that occurs over time and online learning tools generate fine-grained data regarding the temporal aspect of learning. However, the temporal aspect of learning is often neglected while analyzing learner data. According to Chen, Knight, and Wise (2018) temporal study pattern has two features. The first feature is related to the passage of time (how long, how often students engage). The second feature refers to the sequential order in which these activities take place (Molenaar & Järvelä, 2014). Students’ instructional conditions such as learning design influence both of the features. Therefore, the analysis of temporal patterns in the clickstream data tracking student actions is essential to expose the insights of students’ learning processes.

2. eBook Data Analysis in Higher Education

Students’ reading logs previously used to predict their end of year academic performances (Hasnine et al., 2018), developing educational early-warning systems for at-risk students (Akçapınar, Hasnine, Majumdar, Flanagan, & Ogata, 2019a), detecting off-task behaviors during the classroom teaching (Akçapınar, Hasnine, Majumdar, Flanagan, & Ogata, 2019b), modeling students’ level of knowledge (Flanagan, Majumdar, Akçapınar, Wang, & Ogata, 2019), and recently for understanding students’ approaches to learning (Akçapınar, Chen, Majumdar, Flanagan, & Ogata, 2020). In this study, we focus on the temporal aspect of reading logs and try to understand students’ study patterns on a large amount of click-stream data.

3. Method

3.1 Dataset

As the data source, a publicly available dataset collected from the students of Kyushu University was used. Reading logs were collected from the first-year students of the Faculty of Arts and Sciences. Students were registered to the Information Science Course. It was a face-to-face course supported by digital technologies. Weekly course materials related to information science were shared with the students with the help of a digital eBook system (BookRoll).

As mentioned by Ogata et al. (2017) BookRoll is a developed system that allows viewing digital materials used in lectures. It is an online environment that allows teachers to upload content in pdf format which the student can browse anytime and anywhere from a web browser in their personal devices (e.g., laptop, mobile devices, etc.). In BookRoll, there are features like bookmarks, markers, memo functions, which the students can use for learning. Overall data includes 1,914,680 rows click-stream comes from 10 different courses that use the same set of instructional design and learning materials. In each course, an instructor used 8 different learning materials and the overall length of the course is 8 weeks. The overview of the data is given in Table 1. Since all of the courses have the same structure, we merged all the data and treated it as a single course.

Table 1. Overview of the data

<table>
<thead>
<tr>
<th>Course ID</th>
<th>Students</th>
<th>Total Event</th>
<th>Total Reading Session</th>
<th>In Class Reading Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>24a65f29b6</td>
<td>137</td>
<td>164154</td>
<td>2175</td>
<td>1590</td>
</tr>
<tr>
<td>34451e8c77</td>
<td>128</td>
<td>139976</td>
<td>1792</td>
<td>1479</td>
</tr>
<tr>
<td>39a67f80f4</td>
<td>131</td>
<td>207921</td>
<td>2332</td>
<td>1627</td>
</tr>
<tr>
<td>60ab104927</td>
<td>113</td>
<td>248599</td>
<td>2346</td>
<td>1525</td>
</tr>
<tr>
<td>65bb6224af</td>
<td>129</td>
<td>161094</td>
<td>1840</td>
<td>1461</td>
</tr>
<tr>
<td>6b1900c56c</td>
<td>118</td>
<td>263284</td>
<td>2830</td>
<td>1809</td>
</tr>
<tr>
<td>792eaf2c1b</td>
<td>138</td>
<td>190070</td>
<td>2224</td>
<td>1643</td>
</tr>
<tr>
<td>86066cb6a6d</td>
<td>142</td>
<td>199541</td>
<td>2059</td>
<td>1673</td>
</tr>
<tr>
<td>9a683161f5</td>
<td>170</td>
<td>192844</td>
<td>2216</td>
<td>1842</td>
</tr>
<tr>
<td>dbed6c966a</td>
<td>140</td>
<td>147197</td>
<td>1914</td>
<td>1644</td>
</tr>
<tr>
<td>Total</td>
<td>1346</td>
<td>1914680</td>
<td>21728</td>
<td>16293</td>
</tr>
</tbody>
</table>

343
3.2 Data Preprocessing

The provided click-stream data has the following fields: *userid* (anonymized student id), *contentsid* (the id of the eBook that is being read), *operationname* (the action that was done, e.g. open, close, next, previous, jump, add marker, add bookmark, etc.), *pageno* (the current page where the action was performed), *marker* (the reason for the marker added to a page, e.g. important, difficult), *memo_length* (the length of the memo that was written on the page), *devicecode* (the type of device used to view BookRoll, e.g. mobile, pc), and *eventtime* (the timestamp of when the event occurred).

During the preprocessing, we assigned *session id* for each log by separating logs into the reading sessions. While dividing logs into the reading sessions, we considered the *OPEN* event as a starting point of the new reading session. In other words, we coded every sequence of logs as a reading session which starts with the *OPEN* event. By using the information provided in metadata files, we identified material used in each week of the course and assigned lecture id for each course.

**Extracting Study Sequences:** At the final stages of data preprocessing, we extracted students’ learning states for each learning material according to states defined in Table 2. From the click-stream data, it can be understood that all learning materials are uploaded to the system at the beginning of the semester. All activities before the class counted as *Preview* activities and all activities after the class counted as *Review* activities. In other words, if a student opens the material anytime before the class, we considered it as a *Preview* activity. If a student opens the material during the class we considered it as a *Class* activity. If a student opens the material anytime after the class we considered it as a *Review* activity. Based on this information we labeled every single reading session extracted during the preprocessing with one of the states given in Table 2. At the end of this process, each student was given a state for each content. As a result, click-stream data converted into reading sessions and study sequences. Example study sequence for a student could be like this: C1:Preview -> C2:Class -> C3:Inactive -> C4:Class -> C5:Review -> C6:Preview+Class -> C7:Class -> C8:Preview+Class. The description of all types of activities can be seen in Table 2.

<table>
<thead>
<tr>
<th>State</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preview</td>
<td>Activity detected only during the preview period but not during the class and review period</td>
</tr>
<tr>
<td>Review</td>
<td>Activity detected only review period</td>
</tr>
<tr>
<td>Preview+Review</td>
<td>Activity detected during the preview and review period but not during the class</td>
</tr>
<tr>
<td>Class</td>
<td>Activity detected only during the class period</td>
</tr>
<tr>
<td>Preview+Class</td>
<td>Activity detected both during the preview period and the class period</td>
</tr>
<tr>
<td>Class+Review</td>
<td>Activity detected both during the class and review period</td>
</tr>
<tr>
<td>Preview+Class+Review</td>
<td>Activity detected during the preview, class, and review period</td>
</tr>
<tr>
<td>Inactive</td>
<td>No activity detected for a learning material</td>
</tr>
</tbody>
</table>

3.3 Data Analysis

We analyzed sequential data extracted during the preprocessing to understand students’ study patterns across the semester. First, we looked at the general patterns across all data such as the distribution of states for each content. Second, we clustered students based on the similarity of their study patterns. This helped us to see common study patterns followed by the students. Agglomerative Hierarchical Clustering based on Ward’s algorithm (Alexis Gabadinho, Ritschard, Studer, & Müller, 2009) was used to group students with similar study patterns. Optimal matching distance (OM distance) was used as a similarity calculation method. The optimal number of clusters was decided based on the dendrogram of the hierarchical cluster analysis. For labeling the obtained clusters, we compared the visualization of state distribution in each cluster. Data analysis was conducted by the R data mining tool (R Core Team, 2017) with the help of TraMiner (A. Gabadinho, Ritschard, Müller, & Studer, 2011) package.
4. Results

The distribution of students’ reading activities across the semester is illustrated in Figure 1. According to data presented in Figure 1, most of the activities are in Class (60.6%). On the other hand, Preview, Review, and Preview+Review activities are limited in the dataset. Only 1.7% of all activities include these three reading activities. To make output models simpler we merged these activities into the closest activity. Hence, we merged Preview activities with Preview+Class activities, Review activities with Class+Review activities, and Preview+Review activities with Preview+Class+Review activities.

Fig. 1. Distribution of Study Activities.

Agglomerative hierarchical clustering based on Ward’s algorithm was used to group students based on their similar study patterns. The optimum number of clusters was decided to 3 based on the dendrogram of the hierarchical cluster analysis (refer to Fig. 2).

Fig. 2. Dendrogram of the Hierarchical Cluster Analysis.

Fig. 3 shows the distribution of students’ study patterns in each cluster. Each of the horizontal lines in the graph represents one student’s study patterns across all contents. The y-axis shows different learning content used during the course (Content 1 to Content 8). It can be seen from Fig. 3 that students...
in Cluster 1 (n=1046) are mainly active in class and they are 78% of the total students. On the other hand, they do not have much activity before and after class. Students in Cluster 2 (n = 163) are highly active in-class, before the class and also after the class (12%). Students in Cluster 3 (n = 137) view the learning contents during the class in the beginning but they are becoming inactive week by week (10%).

![Cluster 1](image1)

![Cluster 2](image2)

![Cluster 3](image3)

**Figure 3.** Distribution of Study Activities in each Cluster.

### 5. Conclusions

In this study, approximately 2 million click-stream data of 1346 students in the Bookroll platform were analyzed aiming to explore the common study patterns of the students followed during the lectures. The results of the analysis showed that a significant number of the students (~80%) viewed the course materials mostly in the class (e.g. Cluster 1). This result is in accordance with previous studies reporting that students frequently do not read their textbooks before the class (Lieu et al., 2017; Ruscio, 2001). On the other hand, it was observed that only a small number of students (12%) viewed the course materials before in class and after the class (e.g. Cluster 2). According to the findings obtained from a previous study (Akçapınar et al., 2020), it can be speculated that these students can be deep learners or students with high self-regulation skills. However, further studies are needed to test these assumptions and understand individual differences between students in different clusters. It is seen that the students in the last group (e.g. Cluster 3) are active in the class in the first two weeks/contents of the course, but they are mainly inactive in the following weeks/contents. These students might be students who are likely to fail the course and are named as at-risk students in the literature. The obtained results can be used to detect these students in a timely manner.

Although pre-class reading was found correlated with academic performance (Gross et al., 2015; Lieu et al., 2017), our results confirmed that most of the students are not willing to read course contents before and after the class. After-class activities (e.g. formative assessment) are also important for learner-centered systems such as flipped classroom settings (Gilboy, Heinerichs, & Pazzaglia, 2015). Learning analytics can be used for continuous monitoring of students’ study patterns based on reading traces they left on the eBook reader. Hence, timely interventions can be designed to engage students with pre-class reading and after-class activities. It can also be used for facilitating personalization and adaptation to enhance students' learning.
Acknowledgements

This work was partly supported by JSPS Grant-in-Aid for Scientific Research (S) 16H06304 and NEDO Special Innovation Program on AI and Big Data 18102059-0.

References


Assessing young learners’ situational interest in an immersive virtual reality learning environment: the role of epistemic curiosity

Kun-Hung CHENG*

*Graduate Institute of Library and Information Science, Research Center for Humanities and Social Sciences, National Chung Hsing University, Taiwan
* khcheng@dragon.nchu.edu.tw

Abstract: This study aimed to assess young learners’ epistemic curiosity and its role in the learners’ situational interest aroused in an immersive virtual reality (IVR) learning environment. There were 30 elementary school students invited to use an IVR system, namely “Find The ROOT,” for science learning in this study. The results showed that the students’ situational interest could be generally aroused, particularly for the males. Compared with the D-type epistemic curiosity, the students’ I-type epistemic curiosity likely played a more crucial role in their perceptions of situational interest. Moreover, this study supposed that the gender variable might interfere with the relationships between epistemic curiosity and situational interest when young learners were involved in IVR learning activities. Future work was suggested to enlarge the research sample to verify the relationships between epistemic curiosity and situational interest with the consideration of gender issues.

Keywords: epistemic curiosity, situational interest, virtual reality, elementary school

1. Introduction

Interest, a psychological state, was conceptualized as personal interest and situational interest (Krapp et al., 1992). While personal interest refers to a person’s traits in preference of an activity based on past experiences, situational interest is a person’s affective perceptions occurring when he/she interacts with an activity or an environment (Krapp et al., 1992). According to Chen et al. (1999), situational interest consists of six dimensions: total interest, novelty, challenge, attention demand, instant enjoyment, and exploration intention. A recent study reported that a library guide by spherical video-based virtual reality (VR) could enhance university students’ situational interest, particularly for the senses of novelty and challenge (Lin et al., 2019). The affordance of 3D virtual worlds such as Second life for arousing teachers’ situational interest was also addressed by previous studies (i.e., Cho et al., 2015). However, limited work was made for exploring young learners’ situational interest when engaging in fully immersive virtual reality (IVR) environments with head-mounted display (HMD).

Epistemic curiosity is another psychological state that an individual is eager to obtain intellectual knowledge for inherent interest (I-type) or eliminating feelings of informational deprivation (D-type) (Litman, 2008). The role of learners’ epistemic curiosity in their learning behaviors such as information seeking has been addressed by previous studies (e.g., Litman et al., 2005). Since the expedition in IVR learning environments involves behaviors of exploring virtual information, there is a need to examine what role of learners’ traits of epistemic curiosity play in their learning by IVR with considering the possible stimulation of situational interest. In addition, gender differences in users’ perceived presence in VR environments were documented in the literature (e.g., Narciso et al., 2019); hence, the gender issue was also explored in this study.

2. Method

2.1 IVR system

To explore students’ situational interest when engaging in science learning with the aid of IVR technology, this study adopted an IVR system developed by Cheng et al. (2019), namely “Find The
“ROOT,” for plants learning. The learning system was developed based on Oculus platform. Users have to wear an HMD (Gear VR headset was used in this study) along with holding a controller to observe the virtual world and interact with the virtual plants. Students can learn the appearance of plants, the propagation of plants, and the reproduction of plants in the three virtual scenes of the IVR system. For example, when users selected a flower by clicking the trigger button on the controller, a 3D animation for presenting how pollination is achieved by bees would exhibit in front of the users (see Figure 1).

![Figure 1. The screenshots of the IVR learning system (left: selecting a flower, right: 3D animation for pollination).](image)

### 2.2 Participants and research procedure

This study invited 30 elementary school students in 6th grade to participate in the research trial for learning science by IVR. Among these participants, 17 students were male (57%), and 13 students were female (43%). There were 80% of the students reported that they have seen VR application previously, indicating that the effects of technology novelty may not interfere the results of this study.

Before the research trial began, the students were required to responded to a pre-test questionnaire (describe later) for understanding their epistemic curiosity. Besides, each student was individually guided for the usage of the IVR devices (i.e., how to wear the HMD and how to use the controller) by a research assistant. When the student was familiar with the operation of the IVR learning system, the research trial began. The students were informed to freely explore the first two scenes in the virtual plant world. If they felt uncomfortable during the virtual navigation, they were allowed to request for terminating the research trial. As the learning activity finished, the students were required to responded to a post-test questionnaire (describe later) for evaluating their situational interest.

### 2.3 Questionnaires

There were two questionnaires exploited in this study. The epistemic curiosity inventory developed by Litman (2008) was adopted for the pre-test for assessing the students’ psychological traits of epistemic curiosity. The Cronbach’s α values for the two scales were 0.82 (I-type, 5 items) and 0.75 (D-type, 5 items), and overall α value was 0.82. For the post-test, this study adapted the situational interest questionnaire developed by Chen et al. (1999) to fit the context of IVR learning. The Cronbach’s α values for the six scales were 0.93 (total interest, 3 items), 0.70 (novelty, 3 items), 0.80 (challenge, 3 items), 0.89 (attention demand, 3 items), 0.98 (instant enjoyment, 3 items), 0.67 (exploration intention, 3 items), respectively, and the overall α value was 0.74. All the items of the two questionnaires in this study were scored in a 5-point Likert scale, with “1” representing “strongly disagree” and “5” representing “strongly agree.”

### 3. Results

#### 3.1 Perceptions of situational interest

This study firstly conducted repeated ANOVA to compare the students’ scores on the scales of the situational interest. The results presented that there were significant differences among the students’ perceptions of situational interest ($F=93.06$, $p<.001$). Specifically, in the IVR learning activity, the students exhibited strong perceptions of total interest ($M=4.88$, $SD=0.31$), novelty ($M=4.87$, $SD=0.30$), and instant enjoyment ($M=4.84$, $SD=0.42$). Further analyses of gender differences showed that the
males also expressed stronger interest in the three scales than the females did. Although their perceptions of exploration intention \((M=4.74, SD=0.40)\) and attention demand \((M=4.59, SD=0.60)\) were slight weaker than the above three perceptions, the IVR learning system could still engage the students to explore the virtual world and concentrate on the learning tasks. Notably, these students expressed less perceptions of challenge \((M=2.37, SD=1.11)\) than the other perceptions of situational interest, indicating that the IVR learning activity in this study was not a demanding task for the students.

### 3.2 Relations between epistemic curiosity and situational interest

In this study, the young learners exhibited higher epistemic curiosity regarding I-type \((M=4.46, SD=0.51)\) than D-type \((M=3.80, SD=0.65)\). The relationships between the students’ epistemic curiosity and their perceptions of situational interest in the IVR learning activity were further analyzed. As shown in Table 1, the psychological trait of I-type epistemic curiosity was significantly related to the perceptions of exploration intention \((r=0.64, p<.001)\), novelty \((r=0.42, p<.05)\), and total interest \((r=0.44, p<.05)\), indicating that the learners with strong intentions to discover something new for inherent delight (I-type) tended to possess more situational interest, particularly for the behaviors of exploration, sense of novelty, and perceptions of general interest in the IVR learning environment of this study. On the other hand, the D-type epistemic curiosity was not significantly associated with situational interest. Interestingly, when further examining the gender differences in the relations between epistemic curiosity and situational interest, it was found that the females with stronger D-type epistemic curiosity tended to pay more attention to the IVR learning materials \((r=0.56, p<.05)\), implying the gender variable might interfere with the relationships between epistemic curiosity and situational interest.

<table>
<thead>
<tr>
<th></th>
<th>Exploration Intention</th>
<th>Instant Enjoyment</th>
<th>Novelty</th>
<th>Attention Demand</th>
<th>Challenge</th>
<th>Total interest</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-type EC</td>
<td>0.64***</td>
<td>0.33</td>
<td>0.42*</td>
<td>0.35</td>
<td>-0.33</td>
<td>0.44*</td>
</tr>
<tr>
<td>D-type EC</td>
<td>0.29</td>
<td>0.19</td>
<td>0.26</td>
<td>0.02</td>
<td>-0.06</td>
<td>0.26</td>
</tr>
</tbody>
</table>

EC: epistemic curiosity, *p<.05, ***p<.001

### References


Peer Influence, Risk Propensity and Fear of Missing Out in Sharing Misinformation on Social Media during the COVID-19 Pandemic

Ryan EBARDO*, Josephine DE LA CUESTA, Jypzie CATEDRILLA & Santoso WIBOWO

José Rizal University, Philippines
Asia Pacific College, Philippines
Mindanao State University – General Santos, Philippines
Central Queensland University, Australia
*ryan.ebardo@jru.edu

Abstract: Widespread misinformation on social media is high, and this is made worse during a public health crisis. While literature on technological solutions to combat misinformation abounds, limited studies have investigated the psychology behind why misinformation is in rapid diffusion in this digital platform. Using a sample of 209 students, we tested the relationships of peer influence, risk propensity and fear of missing out on the behavioral intention to share misinformation on Facebook during the COVID-19 pandemic. Analysis of the results supported prior literature that peer influence and fear of missing out has a positive influence on the behavioral intention to share misinformation on social media. While risk propensity has a positive influence on sharing misinformation, this cannot be supported at a significant level. We conclude this paper by discussing the implications of our study to academic policies on formal and informal disaster education while highlighting the limitations of the study to provide directions for future scholarly endeavors.

Keywords: misinformation, COVID-19, social media, pandemic, peer influence, FOMO, risk propensity

1. Introduction

Social media has become the foremost source of information and misinformation among digital natives in today’s connected economy. Various social media platforms are highly entrenched in our everyday lives through commerce (Catedrilla, 2017), education (Murire & Cilliers, 2017), health (Tacco, Sanchez, Connolly, & Compeau, 2018) and social relationships (Ham, Lee, Hayes, & Bae, 2019). While social media platforms delivered its promise to establish and nurture social relationships, online social networks have become one of the primary sources of knowledge due to its wealth of information. Users can read, create and share content with their friends and family members with ease (Vranešević, Perić, & Marušić, 2019). The interactive spaces inherent to social media and its phenomenal reach are key factors why it has become an important source of information among its users (Krutka & Carpenter, 2016).

Despite its crucial role in information dissemination for the public good, social media has also become a medium of misinformation. Recent debates in the scholarship have called for the responsible utilization of social media in public communication. Notable among these discourses is the propagation of misinformation in social media to influence political processes (Müller & Schulz, 2019; Tandoc, Lim, & Ling, 2018). The gravity of the misuse of social media platforms is highlighted by calls of several governments to regulate these platforms, to mitigate its abuse and target the spread of deliberate misinformation (Vranešević et al., 2019). Today, the deliberate and undeliberate spread of unverified information on social media has become a public concern that requires attention from various stakeholders (Buchanan & Benson, 2019).

Literature on the misuse of social media has primarily focused on political contexts, and limited research explored the role of these platforms during a pandemic. The scientific orientation of a pandemic requires individuals distributing information to have appropriate training in the appreciation,
interpretation and dissemination of related facts (Hazelton, 2020). While ease and accessibility of social media drive its wide adoption, it has also contributed to the burgeoning challenge in massive misinformation during the COVID-19 pandemic (Frenkel, Alba, & Zhong, 2020). The exponential increase in misinformation has called for different stakeholders such as governments, the World Health Organization (WHO), and social media platforms to adopt a united stand to fight inaccurate information during the pandemic (Guynn, 2020).

In this paper, we provide an overview of influential factors that lead to sharing misinformation in social media during this difficult time. We believe that behavioral intention to share information in social media without adequate verification results to the proliferation of misinformation during a pandemic and merits further scholarly elucidation. Consistent with prior studies, psychological factors play a pivotal role in the behavioral intention to share misinformation. Literature synthesis reveals the lack of theoretical guidance in the study of social media during health crises, with most studies primarily focused on the development of technology solutions to combat misinformation. First, a review of related literature revealed that extant research has been conducted in developed economies despite the fact that misinformation is a key concern in the global south (Silver, 2019). Second, we highlight the importance of social media during a crisis, and we observed that a bulk of related researches primarily focused on investigations related to political processes. Understanding why users interact with social media content during disasters remains under-investigated (Luna & Pennock, 2018). Given the infrequency of a health pandemic, in conducting this research, we identified the influential factors that lead to sharing misinformation during this difficult time. Following the presentation of the current state of the art of social media and misinformation during a pandemic, we discussed our theoretical framework and its corresponding set of hypotheses. This is followed by a discussion of our methodology, the results of our analysis, and conclude with limitations and recommendations.

2. Related Literature and Theoretical Foundations

Wider access to the Internet and the increased popularity of social media translated to unprecedented challenges to the rapid propagation of misinformation online. In the context of this research, misinformation is defined as unverified or inaccurate information that is mostly shared or reshared in social media platforms unintentionally and without due diligence of investigating the veracity of its truthfulness (Wu, Morstatter, Carley, & Liu, 2019). In the study of Bessi and Ferrara (2016), the spread of misinformation appears to be highly influenced by automated social bots during the 2016 United States presidential election. A substantial number of voters were believed to be exposed to misinformation, which they believed to be true and therefore influenced the political landscape in the United States (Allcott & Gentzkow, 2017; Vranesević et al., 2019). In another quantitative study investigating Facebook as a source of news information during the German elections in 2017, the behavior to verify information shared in social media is influenced by prior exposure to misinformation (Müller & Schulz, 2019).

During a crisis, the velocity of the spread of information becomes critical and social media is an ideal platform that supports information exchange necessary for disaster education or safety information dissemination. Using a grounded theory approach, the study of Linlin et al. (2015) revealed that the wide access to social media and its capability to circulate information fast renders it useful during a crisis. However, researchers note that these same reasons can make social media platforms a vehicle to spread misinformation. During pandemic or public health crises, limited studies have investigated the role of social media in the spread of misinformation. Analyzing social media postings regarding Ebola on Twitter, the study of Tran and Lee (2016) revealed the pivotal role of social ties in the propagation of information during a pandemic. Another study focused on Swine Flu and Ebola retrieved and analyzed Twitter posts using several search parameters. Thematic analysis identified fear from both outbreaks as a critical topic that emerged from exaggerated discussions on unverified information on social media (Ahmed, Bath, Shaffi, & Demartini, 2018). In a quantitative study in South Korea on social media and the Middle East Respiratory Syndrome coronavirus (MERS-COV), fear and anger are self-relevant emotions that positively influence social media use highlighting its importance during infectious disease outbreaks (Oh, Lee, & Han, 2020).

Several determinants of behavioral intention have been widely investigated in social media adoption. First, in the Theory of Planned Behavior by Ajzen (1991), subjective norms have been widely tested to positively influence a behavioral intention to perform an act (Hassandoust, Logeswaran, &
Farzaneh Kazerouni, 2011; Lee & Tsai, 2010). In the educational context, social norms are behaviors observed by students on their peers who are mostly their classmates. Peers are within the social ties of students in their social media accounts, and sharing misinformation within their own networks may be perceived to be typical, with utter disregard for appropriate information verification. As such, we hypothesize that:

H1: Peer influence positively influences behavioral intention to share misinformation

Second, social media interactions involve varying levels of risk. As a public space, people interact with strangers, share photos, post and repost verified and unverified information online. Prior literature revealed that an individual with a high level of risk propensity is more likely to share misinformation online (Kooihkamali & Sidorova, 2017). As such, we hypothesize:

H2: Risk propensity positively influences behavioral intention to share misinformation

Social technologies allow individuals to belong to an online community where they interact with family members, friends and peers. The behavior demonstrated by people in a social network exerts pressure on an individual to behave the same as others in order to minimize isolation, a concept known as Fear of Missing Out or FOMO. On the use of social media and deviant Internet behavior, the psychological effect of FOMO has been found to have a positive influence on problematic social media use (Reyes et al., 2018). As such, we hypothesize:

H3: Fear of Missing Out positively influences behavioral intention to share misinformation

In the context of this study, we hypothesize that Peer Influence, Risk Propensity and Fear of Missing Out are psychological determinants in the behavioral intention to share misinformation on social media as summarized in Figure 1 – Theoretical Framework:

![Diagram](image)

*Figure 1. Theoretical Framework and Hypotheses*

3. Methods and Procedures

To validate the structural model, we utilized a scale consisting of twenty-four (24) items adopted from prior literature. We utilized three (3) items for Peer Influence from Chang (2014), three (3) items for Risk Propensity and five (5) items for intention from Kooihkamali and Sidorova (2017), and ten (10) items were adopted from Abel, Buff and Burr (2016). To contextualize the scale, we updated some terms to fit within the domain of our study. To test the validity and reliability of the scale, we applied a Partial Least Square Algorithm using SmartPLS. Indicators that did not meet the minimum threshold were deleted until the minimum values for Cronbach Alpha (0.70), Average Variance extracted or AVE (0.50) and Composite Reliability or CR (0.70) were obtained. The values are shown in Table 1 – Scale Validation Results:
Table 1. 
Scale Validation Results

<table>
<thead>
<tr>
<th>Construct</th>
<th>Cronbach’s Alpha</th>
<th>Composite Reliability</th>
<th>Average Variance Extracted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peer Influence</td>
<td>0.945</td>
<td>0.948</td>
<td>0.858</td>
</tr>
<tr>
<td>Risk Propensity</td>
<td>0.805</td>
<td>0.885</td>
<td>0.720</td>
</tr>
<tr>
<td>FOMO</td>
<td>0.858</td>
<td>0.886</td>
<td>0.610</td>
</tr>
<tr>
<td>Intention</td>
<td>0.891</td>
<td>0.906</td>
<td>0.710</td>
</tr>
</tbody>
</table>

Given the manner of our data collection and theoretical operationalization, we tested for the existence of common method variance which may result from the way the items in the questionnaire are presented, the use of the same respondents for both dependent and independent variables, and the presence of social desirability bias (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). To address this potential issue, we conducted a full collinearity test and extracted the variance inflation factor values or VIFs (Kock, 2015). The test yielded values that are less than 3.3, where the values of 1.067, 1.087 and 1.021 for peer influence, risk propensity and fear of missing out, respectively, indicate that the study does not suffer from common method bias (Ifinedo, 2017; Kock, 2015).

To further establish the validity of the survey instrument, we tested for discriminant validity through a Fornell-Larcker criterion test. Examination of the results is a prerequisite to ensure that each construct is able to depict the measure it was designed to represent (Hair, Hult, Ringle, & Sarstedt, 2014). As shown in Table 2 – Fornell-Larcker Test, the square root of the AVEs for each construct is higher than the other inter-construct correlation values, therefore, establishing discriminant validity (Nelson, Verhagen, & Noordzij, 2016; Yang, Yu, Zo, & Choi, 2016).

Table 2. 
Fornell-Larcker Test Results

<table>
<thead>
<tr>
<th></th>
<th>FOMO</th>
<th>Intention</th>
<th>Peer Influence</th>
<th>Risk Propensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOMO</td>
<td>0.781</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intention</td>
<td>0.263</td>
<td></td>
<td>0.843</td>
<td></td>
</tr>
<tr>
<td>Peer Influence</td>
<td>0.101</td>
<td>0.249</td>
<td>0.927</td>
<td></td>
</tr>
<tr>
<td>Risk Propensity</td>
<td>0.034</td>
<td>0.173</td>
<td>0.066</td>
<td>0.848</td>
</tr>
</tbody>
</table>

Recent arguments against the use of the Fornell-Larcker criterion test in establishing discriminant validity for variance-based empirical research necessitates further evaluation of the survey instrument to establish sound confirmatory factor analysis procedure (Ab Hamid, Sami, & Mohmad Sidek, 2017; Henseler, Ringle, & Sarstedt, 2014). To ensure that the latent variables do not suffer from multicollinearity issues, we tested for discriminant validity using the Heterotrait-Monotrait values. As shown in Table 3 – Heterotrait-Monotrait Test, all values are below 0.90, which further supports that the instrument demonstrates strong discriminant validity among its operationalized constructs (Ab Hamid et al., 2017).

Table 3. 
Heterotrait-Monotrait Test Results

<table>
<thead>
<tr>
<th></th>
<th>FOMO</th>
<th>Intention</th>
<th>Peer Influence</th>
<th>Risk Propensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOMO</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intention</td>
<td>0.225</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peer Influence</td>
<td>0.124</td>
<td>.097</td>
<td>0.184</td>
<td>0.156</td>
</tr>
<tr>
<td>Risk Propensity</td>
<td>0.153</td>
<td>.161</td>
<td>0.184</td>
<td>0.156</td>
</tr>
</tbody>
</table>

To test our hypotheses, we deployed an online survey from March 15 to April 15 during the COVID - 19 lockdown to students from two universities in the Philippines. Participation is voluntary.
and the participants gave informed consent. A total of 209 students provided valid responses. All students are enrolled in various undergraduate programs in colleges and universities with active Facebook accounts. To produce better approximation and due to the small sample size, a bootstrapping technique (Schmidheiny, 2014) was applied to the responses and the results are shown in Table 4 – Structural Model Test Results:

<table>
<thead>
<tr>
<th>HYPOTHESIS</th>
<th>SD</th>
<th>T STATISTICS</th>
<th>P Values</th>
<th>DECISION</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1: Peer influence positively influence behavioral intention to share misinformation</td>
<td>0.079</td>
<td>7.534</td>
<td>0.000</td>
<td>Accept</td>
</tr>
<tr>
<td>H2: Risk propensity positively influence behavioral intention to share misinformation</td>
<td>0.136</td>
<td>0.689</td>
<td>0.491</td>
<td>Reject</td>
</tr>
<tr>
<td>H3: Fear of Missing Out positively influence behavioral intention to share misinformation</td>
<td>0.077</td>
<td>2.770</td>
<td>0.006</td>
<td>Accept</td>
</tr>
</tbody>
</table>

4. Discussion of Results

Based on the results, we confirm that Peer Influence (H1) and Fear of Missing Out (H3) positively influence the behavioral intention to share misinformation due to their T-Statistics scores of 7.543 and 2.770, respectively. As they are higher than the minimum score, we can support both hypotheses at a significant level (Hair et al., 2014). These findings are consistent with prior literature confirming the role of peers in deviant online behaviors and the growing influence of social media in the instigation of fear to be left out among younger adults (Burnett, Enyeart Smith, & Wessel, 2016; Rice & Staffo, 2012). Students consider people in academic institutions as peers and the likelihood of them being in the ‘Friends’ list is high. In the study of Riemenschneider, Leonard and Manly (2011), students’ unethical behaviors are highly influenced by normative beliefs drawn from how people they consider important behave. Peer influence is formed with the approval of significant persons prior to performing a specific act and in the context of social media, these peers are within the network with whom they have close ties (Khan & Idris, 2019). In Facebook, content posted by people within a students’ social network can influence the decision process to re-share such information. In a recent study, students intimated that they would re-share unverified content if they see that these news information were also shared or liked by other people in their own social networks (Oh et al., 2020).

Like the influence of social networks, the fear of being left out can also influence the decision to share misinformation. Individuals who have high FOMO tend to behave like most of the people in a group to satisfy their urge to socially belong. In addition, prior literature have demonstrated FOMO as a strong predictor of deviant behavior among young adults in social media (van Rooij, Lo Coco, De Marez, Franchina, & Abeele, 2018). Prior literature linked FOMO to problematic online behavior such as social media addiction (van Rooij et al., 2018). During the community lockdown in the Philippines, educational institutions were forced to close its operations and shift to online platforms to support learning (Bagayas, 2020). Such transition affords students more time to spend online and a recent study points to high social media exposure among individuals during the COVID-19 pandemic has caused mental problems (Wang et al., 2020). While the T-statistics value of H2 is positive, this cannot be supported at a significant level. Our findings support the study of Buchanan & Benson (2019), which did not find a positive influence between risk propensity and the behavior to spread unverified information. A possible explanation is that sharing or reposting of unverified information is considered a low-risk behavior. During the COVID-19 pandemic, the government enacted the Bayanihan to Heal as One Act or Republic Act 11469 that includes provisions that sanction individuals who spread fake news and misinformation online (Sawada&gan, 2020).
5. Conclusion and Recommendations

In conclusion, we confirm that peer influence and fear of missing out (FOMO) can positively influence the dissemination of misinformation on social media during a crisis. We contribute to recent calls in social media research to go beyond technology by approaching the challenge of misinformation during disasters through understanding the psychology behind this systemic online behavior (Elbanna, Bunker, Levine, & Sleigh, 2019; Luna & Pennock, 2018). As more people harness information online, social media becomes important during difficult times, and the role of the academe in formal and informal education is equally vital in mitigating the spread of misinformation.

Central to discourses on disaster education is the function of the academe in the management of misinformation on social media. Peer influence or social pressure on students can come from different sources such as classmates, close friends and family members. In this regard, it will be timely to update the present curricula on disaster education to include social media ethics to indoctrinate appropriate online ethical behavior among students (Harris & Lang, 2011). While technology supports formal and informal learning, it ushers ethical dilemmas within and outside the hallways of the classroom (Gutierrez & Padagas, 2019; Trapero, 2018). Teachers and parents can exert ethical influence on students and academic institutions can find innovative activities where both can cooperate to further understand students’ behavior online (Bagnall, Skipper, & Fox, 2019). Research indicates that the fear of missing out is a direct result of excessive social media usage. Intervention such as self-monitoring of social media use can be effective in addressing this behavior (Dogan et al., 2019). Institutions can promote the benefits of social media for educational purposes to engage students in meaningful activities rather than actively immersing themselves in misinformation. Additionally, academic institutions can go beyond the traditional hallways of the classroom by engaging in informal education through social media. This can be accomplished by involving healthcare experts and authorities to participate as informal educators in academic social media communities to provide content to support disaster education measures during a health pandemic (Feng, Hossain, & Paton, 2018).

Our findings should be interpreted with its limitations in mind and its corresponding research opportunities. First, the sampling technique and small sample size might invite questions on its generalizability and therefore we encourage future research to test our study on a larger population and a longitudinal technique to investigate the influence of the constructs after the crisis. Second, the study was conducted in the Philippines and therefore it will be interesting to compare these results with another cultural context. Lastly, while we have confirmed two hypotheses and rejected one, we utilized a methodology that can be improved to provide deeper insights to the results of this inquiry. Future research can extend this study by testing the moderating effects of age and gender as well as applying qualitative techniques to further elucidate the results of this paper. Another opportunity for future research is the application of modern methodology such as netnography to analyze comments on unverified information by users, to capture how they influence students to share such content.

References


!CFE<22CA6?E6C  &2CE:4:A2E@CJ"62C?:?8E9C@F89)@4:2=#65:2@H2?5-9J)@4:2=)EF5:6D
5F42E@CD+D6*H:EE6C76<.587:*:A;;=.;26$.,167470A*6-$.*,1.:-=,*<276   R
"66 #  *D2: *( 
 -92E C:G6D &6@A=6 E@ @?E:?F6 E@ &=2J %?=:?6 2>6D ? IE6?D:@? @7
*649?@=@8J#@56=2?5*96@CJ@7&=2??65692G:@C6<.:6*<276*47=:6*47/=5*6758=<.:6<.:*,<276
 R 9EEAD 5@:@C8  

 
":?=:?F2?8/)E2C3:C5!%C2?5#)E2?6<)&656CD6?* @??64E65E9C@F894C:D:D
>@E:@?2=AC@I:>:EJ2?5E96DAC625@7>:D:?7@C>2E:@?@?=:?6#'  !:7,..-260;7/<1.  
6<.:6*<276*4 76/.:.6,. 76 758=<.:#=887:<.- 778.:*<2>. '7:3 *6- #7,2*4 758=<260 R 
9EEAD 5@:@C8      
2?5 C6D62C49 286?52 6<.:6*<276*4 7=:6*4 7/ 2;*;<.: "2;3 ".-=,<276  R
9EEAD 5@:@C8   ;:;5CC   
#N==6C&)49F=K 2463@@<@C2<63@@<@HFD6CDVA6C46AE:@?D@7U72<6?6HDV2C6C6=2E65E@
9EEAD 5@:@C8 
   
#FC:C6%*:==:6CD" )@4:2=>65:225@AE:@?2>@?8=64EFC6CD2E2EC25:E:@?2=F?:G6CD:EJ:?2DE6C?
2A6 &C@G:?46 @7 )@FE9 7C:42 # 7=:6*4 7/ 6/7:5*<276 *6*0.5.6<   R
9EEAD 5@:@C8 
D2;:>G :  
$6=D@?   ,6C9286? *  $@@C5K:; # "   62=E9 6>A@H6C>6?E E9C@F89 24E:G:EJ EC24<6CD ?
6>A:C:42= D>2CE HC:DE32?5 DEF5J 758=<.:; 26 =5*6 .1*>27:  
R 
9EEAD 5@:@C8   ;493   
%9 )  "66 ) /  2?  
 *96 7764ED @7 )@4:2= #65:2 +D6 @? &C6G6?E:G6 692G:@CD 5FC:?8
?764E:@FD:D62D6%FE3C62<D*96#65:2E:?8(@=6@7)6=7C6=6G2?E>@E:@?D2?5&F3=:4(:D<&6C46AE:@?
.*4<1755=62,*<276  R 9EEAD 5@:@C8  


   
&@5D2<@77&##24!6?K:6)"66 /&@5D2<@77$&
@>>@?#6E9@5:2D6D:?692G:@C2=
(6D62C49C:E:42=(6G:6H@7E96":E6C2EFC62?5(64@>>6?565(6>65:6D7=:6*47/8842.-!;A,17470A
R 9EEAD 5@:@C8  
 
(6J6D#)#2C2D:82? &@?K2=6D '6C?2?56K!"##65:@D#%2JF3:E(
% 62C@7>:DD:?8@FE2?5:ED=:?<H:E9D@4:2=>65:22?5AC@3=6>2E:4:?E6C?6EFD62>@?87:=:A:?@D
7:<15.:2,*67=:6*47/!;A,17470A   R 
(:46#)E277@#
#**$&"()#$"($*(%+*%(/%+())$3J
$$   !"( )F3>:EE65 :? A2CE:2= 7F=7:==>6?E @7 E96 C6BF:C6>6?ED 7@C E96 568C66 @7 @4E@C @7
&9:=@D@A9J:?E96?E6C5:D4:A=:?2CJ&C@8C2>D@75F42E:@?2="6256CD9:A&@=:4J2?5*649?@=@8J)EF5
(:6>6?D49?6:56C  ! "6@?2C5 " $ !  #2?=J * ) 
 )EF56?EDV 6E9:42= 564:D:@?>2<:?8 :? 2?
:?7@C>2E:@?E649?@=@8J4@?E6IEE96@CJ@7A=2??653692G:@C2AAC@2497=:6*47/6/7:5*<276#A;<.5;
-=,*<276
 
R 
(6EC:6G65
7C@>
9EEA D62C49AC@BF6DE4@>AC@IJ8C6?@3=6
6>4@> 3FD4@== 5@4G:6H    7F==E6IE&    &' 244@F?E:5 
)2H25;22?

7C@>9EEAD HHH3H@C=5@?=:?64@> D:>A=:7J:?8E9632J2?:92?E@962=2D@?624E 
)49>:596:?J !   =FDE6C:?8 :? E96 ":?62C #@56= #17:< =2-.-; <7 2,:7.,7675.<:2,;  %62>.:;2<*.<
*;.4 R 9EEAD 5@:@C8   ;86?9@DAADJ49
  
):=G6C" #:D:?7@C>2E:@?2?5762CD23@FE:ED:>A24E2C6A6CG2D:G6:? 6>6C8:?864@?@>:6D(6EC:6G65
AC:= 
7C@>9EEAD HHHA6HC6D62C49@C8 724EE2?<
 
>:D:?7@C>2E:@?2?5762CD23@FE:ED
:>A24E2C6A6CG2D:G6:? 6>6C8:?864@?@>:6D 
*244@#56))2?496K%@??@==J(@>A62F ?6I2>:?2E:@?@7E962?E64656?ED@7ECFDE
:? 72463@@< @?=:?6 962=E9 4@>>F?:E:6D <1 =:78.*6 76/.:.6,. 76 6/7:5*<276 #A;<.5; .A76-
202<2B*<276*,.<;7/#7,27$.,162,*41*60.#  
*2?5@4":>0-":?8( 67:?:?8S2<6$6HDTEJA@=@8J@7D49@=2C=J567:?:E:@?D202<*4
7=:6*42;5  R  9EEAD 5@:@C8  
      
*C2?*"66! +?56CDE2?5:?84:E:K6?C624E:@?D2?53@=2C6=2E65:?7@C>2E:@?AC@A282E:@?@?D@4:2=
>65:2 !:7,..-260; 7/ <1.    6<.:6*<276*4 76/.:.6,. 76 ->*6,.; 26 #7,2*4 .<?7:3;
6*4A;2;*6-26260#    R 9EEAD 5@:@C8   )%$# 

*C2A6C@     F8>6?E65 C62=:EJ :? :??@G2E:?8 A6528@8J E9:42= :DDF6D @? A6CDF2D:G6 E649?@=@8:6D
   <1 6<.:6*<276*4 76/.:.6,. 76 758=<.:; 26 -=,*<276 *26 76/.:.6,. !:7,..-260;
646>36CR
G2?(@@:; "@@4@6#2C6K"C2?49:?2,366=6#,2?56? 62C@7>:DD:?8@FE2D
2AC65:4E@C@7AC@3=6>2E:4D@4:2=>65:2FD62?5A9F33:?83692G:@C2>@?87=6>:D925@=6D46?ED6<.:6*<276*4
7=:6*47/6>2:765.6<*4".;.*:,1*6-!=+42,.*4<1  9EEAD 5@:@C8   :;6CA9 

,C2?6Y6G:Z*&6C:Z$#2CFY:Z* &6C46AE:@?@7)@4:2=#65:22D2)@FC46@7(6=6G2?E?7@C>2E:@?
)*0:.+ 6<.:6*<276*4 ".>2.? 7/ ,76752,; *6- =;26.;;  
R  9EEAD 5@:@C8   K:C63
 

358




The Influence of Augmented Reality Embedding Cognitive Scaffolds on Elementary Students’ Scientific Learning

Xinyue JIAOa, Zifeng LIUa, Su CAIab, c *

a VR/AR+Education Lab, School of Educational Technology, Faculty of Education, Beijing Normal University, China
b Beijing Advanced Innovation Center for Future Education, Beijing Normal University, China
c Research Institution of Science Education, Beijing Normal University, China
* caisu@bnu.edu.cn

Abstract: Augmented Reality (AR) can enhance students’ learning performance by visualizing abstract concepts. However, most AR applications in the classroom just simply show virtual-real combined scenes without deeply integrated with learning materials. In that case, students would always have a lot of difficulties to construct knowledge by themselves. This study developed an AR application embedding cognitive scaffolds to help students better construct the scientific knowledge and explore the impact of those cognitive scaffolds on students learning. A quasi-experimental method was used by dividing 42 students into three groups randomly, to investigate students’ scientific achievements, learning experience and cognitive load. Students in experimental group 1 learned with AR without the support of cognitive scaffolds, and students in experimental group 2 learned with AR embedding cognitive scaffolds, while students in the control group learned with traditional method. Semi-structured interview was conducted after class. It is found that teaching with AR technology could significantly improve students’ scientific achievements and learning experience. Compared with AR application without the support of cognitive scaffolds, AR with cognitive scaffolds has no significant impact on students’ scientific achievements, learning experience, and cognitive load. Students in both experimental groups have better learning experience and lower cognitive load. The interview results revealed four advantages of AR technology: promoting student’s cognition, having positive learning emotions, improving students’ ability of hands-on and observation, convenient to use. And students’ different attitudes towards the cognitive scaffolds were found in the interview, which might explain why the effects of cognitive scaffolds are not very significant.

Keywords: Augmented Reality, Science Education, Cognitive Scaffolds

1. Introduction

Augmented Reality (AR) has a powerful potential in education (Freeman, Adams Becker, & Cummins, 2017). As an extension of Virtual Reality (VR) technology, AR technology has the unique advantage, which can break the boundary between virtual space and real space (Milgram, Takemura, Utsumi, & Kishino, 1994). It brings new opportunities for education by helping students better understand abstract concepts, making up for the shortcomings of multimedia teaching in the past. In recent years, AR has been widely applied to science education, especially in improving students’ spatial abilities, conceptual understanding and inquiry ability (Arici, Yildirim, Caliklar, & Yilmaz, 2019).

However, several research focuses on technology and ignores instructional design when applying AR in the classrooms, which causes many difficulties for students (De Jong & Van Joolingen, 1998). Therefore, some researchers have proposed scaffoldings to overcome these difficulties (Rutten, Joolingen, & Veen, 2012). A few studies showed that embedding cognitive scaffolds in AR is helpful (Wu, Hwang, Yang, & Chen, 2017). But according to the cognitive load theory, the scaffolds cannot help students’ learning (Yoon, Elinich, Wang, Steinmeier, & Tucker, 2012) even increase the cognitive load of students (Ibili, 2019). Therefore, we developed an AR application integrated the
supports of cognitive scaffoldings to explore the influence of the cognitive scaffolds in AR on students’ scientific learning, and to investigate students’ views on AR and cognitive scaffoldings in it.

2. Literature Review

2.1 AR in Scientific Learning

Since the subject of science is mainly based on abstract content such as text symbols, students usually cannot really grasp some abstract concepts in scientific knowledge. AR provides a kind of ways to solve such difficulties. The character about combining virtual objects and real world makes it have a strong ability to establish situation. Therefore, AR could help students understand abstract concepts, enhance the learning experience, improve learning achievements and have a more positive attitude (Ibáñez & Delgado-Kloos, 2018), playing an important role in the field of science education.

Optical knowledge is one of the important concepts in science. In the traditional science class, previous studies revealed that students usually misunderstand the concepts in optics, such as: confusing the two concepts of sight and light (Andersson & Bach, 2005); confusing the color of the light itself with the color of the light reflected on the object. These misconceptions indicated that learners lack understanding of the presence of light, and the connection between colors of light and the colors of observed objects. In order to solve the above difficulties, researchers suggested using technology to help students understand the optics knowledge, such as simulation, game-based learning (Hvannberg, Law, & Hallodrsdottir, 2018). A few researchers applied AR technology to teaching optical experiments for middle school students, such as convex imaging experiment, single-slit experiment and double-slit experiment (Cai, Chiang, & Wang, 2013; Niu, Xu, Cheng, & Cai, 2018; Wang, Zhang, Xue, & Cai, 2018), which confirmed the positive role of AR technology in assisting students to learn optics. It is necessary for educational researchers to take advantage of AR to develop relevant educational applications to help pupils’ learning optics knowledge mentioned above.

2.2 Scaffolds in AR Learning Environments

When applying AR applications to a real teaching environment, three factors of content, technology, and pedagogy should be taken into the consideration comprehensively (Bidarra & Rusman, 2016). However, some studies of AR technology applying in the classrooms usually ignored instructional design well integrated with AR. For example, some studies showed that teaching with AR for inquiry activities had not improved the learning achievements (Kirschner, Sweller, & Clark, 2006). Therefore, after discussing and studying on these issues, some researchers proposed that scaffoldings maybe overcome these problems (Rutten et al., 2012). The scaffoldings are temporary supports and assistance provided by teachers or simulations. There are many types of scaffoldings, such as cognitive, meta-cognitive, motivational, collaborative and knowledge-building scaffoldings (Tsai & Huang, 2014). Cognitive scaffoldings refer to the support, prompts, suggestions and assistance about the learning content, and strategies related to the problem solving. In recent years, researchers have considered to using computers to provide effective scaffoldings, such as online teaching platforms, educational games (Hwang, Sung, Hung, Yang, & Huang, 2013). A few studies concluded that cognitive scaffoldings in AR would support a better metacognitive process and reduce students’ cognitive load (Ibanez, Di-Serio, Villaran-Molina, & Delgado-Kloos, 2016; Wu et al., 2017).

However, some studies results argued that the embedding scaffoldings into AR is not very helpful for students’ learning. For example, Yoon’s research indicated that some cognitive scaffoldings may not be necessary to construct general concepts except for digital augmentation (Yoon et al., 2012). Although AR has a certain potential for reducing the cognitive load of students (Cheng, 2016), there are also some studies showed that too many elements (such as text, prompts, etc.) in the AR system would distract students’ attention and increase students’ cognitive load, by increasing extraneous cognitive load (Ibili, 2019). In order to reduce cognitive load, it is necessary to remove interesting but irrelevant multimedia content and reduce the number of interactive elements (Mayer & Pilegaard, 2014). It is not clear that how to integrate learning cognitive scaffoldings with AR tools in the classrooms.
2.3 Research questions

The purpose of this study is to investigate the effects of AR applications and the cognitive scaffolds on students’ performance in scientific learning. Further, we try to analyze the advantages and disadvantages of AR and cognitive scaffolds in it through interviewing teachers and students. The research questions of this research are as follows:

RQ1: Compared with traditional methods, can the AR learning method improve students’ scientific achievements, learning experience, and reducing cognitive load?
RQ2: Compared with AR learning approach without the support of cognitive scaffolds, can the AR embedded cognitive scaffolds improve students’ scientific achievements, learning experience, and reducing cognitive load further?
RQ3: What do students and teachers think about using AR integrated with cognitive scaffolds for learning/teaching?

3. Methods

3.1 Participants and Procedure

Because of COVID-19, the present study was mainly conducted through online teaching. A total of 42 sixth grade elementary school students participated in this experiment. They are from different provinces in China, including 22 boys and 21 girls, with an average age of 11 years (SD = 1.408). They were divided randomly into two experimental groups (EG) (n = 14 per group) and a control group (CG) (n = 14). All students did not learn the related scientific knowledge of optics before.

The experimental procedure is presented in Figure 1. A quasi-experimental study design was used for this study. Before the course begins, there was a brief introduction concerning the basic concepts about light and the basic learning activities. After that, all groups of students were asked to take a pre-test to assess their prior knowledge of the scientific knowledge of the three primary colors of light. The classes were conducted through online video conferencing software. All participants used their parents’ phone or tablet computer at home under the instruction of the teacher. The learning activities included situational introduction, inquiry, summary and practice, lasting for 45 minutes in total approximately. The students in the experimental group 1 learned with the AR application without cognitive scaffolds, while students in experimental group 2 learned with AR application integrated with cognitive scaffolds. On the other hand, students in the control group learned with traditional methods. The learning content and the teacher were same in all three groups except for experimental condition. After the learning activities, students were told to complete the post-test and post-questionnaires of
cognitive load and learning experience. Finally, five students were recruited from two experimental groups to take part in a semi-structured interview to probe their opinions during learning with the AR application. The teacher was also interviewed.

3.2 Experimental Materials: AR applications

AR applications were developed by Unity3D and Vuforia on Windows 10 system, and then they were packaged into teaching software suitable for Android system. A total of two versions of AR applications were developed (Figure 2). Version one was used for experimental group 1 and there were no scaffoldings embedded into it; version two was used for experimental group 2 and cognitive scaffoldings were embedded into the AR system. Cognitive Scaffoldings in AR application included text highlighting, learning tips, feedback on student responses, etc. For example, we displayed and highlighted the specific text on the interface, and set up a "Learning tips" button. A learning cue for this scenario will be shown when students click on "Learning tips" button. Apart from the scaffoldings, the interface and activities in the two versions were identical.

![Figure 2. Two versions of AR applications. Left is version one (without scaffoldings), right is version two (with cognitive scaffoldings).](image)

The AR applications were developed around the scientific concept of "the three primary colors of light" during the primary school period. There are 3 main scenes. The first scene is situational introduction. When scanning the special tracker cards called "marker", a virtual dancing girl will show up. Students can observe the color of the stage lights and the color of the girl’s skirt, then the concept of the three primary colors of light were introduced; the second scene is about inquiry activity. Students can drag the light by finger and change the intensity of the light (click "+" or "+") to explore the mixing law of the three primary colors of light, and reach a conclusion finally; the third scene is about summary and practice activity. A total of three questions are proposed on the interface for students to practice and apply the mixing law of the three primary colors of light into real life.

3.3 Measuring Tools

Scientific knowledge test: The test was designed by a science teacher to evaluate the conceptual understanding and application of students’ scientific knowledge of the three primary colors of light. Both the pre-test and post-test had 10 items, including 6 multiple-choice items, 3 yes-or-no items and 1 open-ended question. The full score was 100 points. Its Cronbach’s α values was 0.73.

Cognitive load questionnaire: This questionnaire used a scale designed by Paas (Paas & Merriënboer, 1994). It consisted of two dimensions of mental load and mental effort. All items used a 9-point Likert rating scheme. In mental load dimension 1 meant "very easy", 9 meant "very difficult", and in mental effort dimension 1 meant "least effort" and 9 meant "maximum effort". Its Cronbach’s α was 0.74.

Learning experience questionnaire: This questionnaire was adopted from the scale developed by Stull (Stull, Fiorella, Gainer, & Mayer, 2018). It included three dimensions of learning motivation, interactive feeling and learning investment. There were 10 items in total, and the 7-point Likert scale
was used. Higher scores indicated better learning experience. The Cronbach’s α value of this questionnaire was 0.90, showing a high reliability.

The interview questions were modified from the questions developed by Hwang, Yang, Tsai, and Yang (Hwang, Yang, Tsai, & Yang, 2009). We used an audio recorder to record the interview data. The questions included but were not limited to the following:

- Do you remember the parts of the AR class? What did you learn separately?
- Which part of the course do you like best? Or what activities do you think are particularly interesting in this class? Why?
- How is this class different from the science class you have taken before? Why?
- What are the advantages of AR application? What are the disadvantages?
- What abilities did you improve in this class? Or what knowledge did you learn?
- Do you like the tips and feedback in AR application? What do you think of them?

4. Results

4.1 Scientific Knowledge Test

A paired-sample t-test was conducted to analyze the difference between pre-test and post-test scientific knowledge scores of the three groups. The results are shown in Table 1. The post-test scores of the three groups of students were significantly higher than the pre-test scores. This means the students’ performance has significant improvements after this science class.

Table 1. The Paired t-Test results for Pre-Test and Post-Test Score Variables of Three Groups

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Pre-test</th>
<th>Post-test</th>
<th>MD</th>
<th>t</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>CG</td>
<td>14</td>
<td>36.42 (17.80)</td>
<td>66.07 (18.10)</td>
<td>29.64</td>
<td>4.04**</td>
<td>13</td>
</tr>
<tr>
<td>EG-1(AR)</td>
<td>14</td>
<td>29.64 (19.16)</td>
<td>81.43 (16.69)</td>
<td>51.78</td>
<td>8.78***</td>
<td></td>
</tr>
<tr>
<td>EG-2 (AR + scaffoldings)</td>
<td>14</td>
<td>37.5 (17.18)</td>
<td>90.00 (14.41)</td>
<td>52.50</td>
<td>9.23***</td>
<td></td>
</tr>
</tbody>
</table>

** p < .01, *** p < .001

In order to compare the differences of the post-test scores between the three groups, ANCOVA was used by excluding the interference of the three groups’ prior knowledge. The pre-test scores were used as covariate, and the post-test scores were used as a dependent variable. Before the analyze of ANCOVA, a homogeneity test was firstly performed. The assumptions of homogeneity of regression was satisfied ($F(2,38) = 0.146, p > .05$), then ANCOVA was performed. The results are shown in Table 2. It can be seen that excluding the effect of pre-test scores, students’ post-test scores were significantly different between the three groups ($F(2, 38) = 7.426, p = .002$). The pairwise comparison results showed that the post-test scores of the students in the EG-1 (Adjusted $M = 81.68, SD = 16.69$) was significantly higher than the CG (Adjusted $M = 65.97, SD = 18.10$), indicating that teaching with AR was better than traditional methods significantly. In addition, although the post-test scores of EG-2 (Adjusted $M = 89.84, SD = 14.41$) were higher than those of EG-1 (Adjusted $M = 81.68, SD = 16.69$), there was no significant difference between the EG-1 and EG-2.

Table 2. The ANCOVA Results of the Students’ Scientific Knowledge Test

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>M (SD)</th>
<th>Ad M</th>
<th>SE</th>
<th>F</th>
<th>η²</th>
<th>Post-hoc</th>
</tr>
</thead>
<tbody>
<tr>
<td>CG</td>
<td>14</td>
<td>66.07 (18.10)</td>
<td>65.97</td>
<td>4.46</td>
<td>7.42**</td>
<td>.281</td>
<td>EG1&gt;CG*</td>
</tr>
<tr>
<td>EG-1(AR)</td>
<td>14</td>
<td>81.43 (16.69)</td>
<td>81.68</td>
<td>4.51</td>
<td></td>
<td></td>
<td>EG2&gt;CG***</td>
</tr>
<tr>
<td>EG-2 (AR + scaffoldings)</td>
<td>14</td>
<td>90.00 (14.41)</td>
<td>89.84</td>
<td>4.74</td>
<td></td>
<td></td>
<td>EG2&gt;EG1</td>
</tr>
</tbody>
</table>

* p < .05, ** p < .01, *** p < .001

4.2 Cognitive Load

As shown in Table 3, ANOVA was conducted to examine the difference of cognitive load between three groups. In the dimension of mental effort, there was no significant difference among the
three groups ($F(2, 38) = 2.909, p > .05$); in the dimension of mental load, the mean scores of the control group and the experimental group were 3.71 ($SD = 1.72$), 3.36 ($SD = 2.49$), 2.79 ($SD = 1.62$), and the difference among the three groups was not statistically significant ($F(2, 38) = 0.776, p > .05$). It is worth noting that in the dimension of mental load, the values of the three groups were all lower than the average of 5, which meant that the students of the three groups all perceived a lower cognitive load in the learning activities.

### Table 3. The ANOVA Results of the Students’ Cognitive Load

<table>
<thead>
<tr>
<th>Variable</th>
<th>Group</th>
<th>N</th>
<th>M (SD)</th>
<th>SE</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mental effort</td>
<td>CG</td>
<td>14</td>
<td>6.43 (1.50)</td>
<td>0.40</td>
<td>2.909</td>
</tr>
<tr>
<td></td>
<td>EG-1 (AR)</td>
<td>14</td>
<td>7.21 (2.32)</td>
<td>0.62</td>
<td></td>
</tr>
<tr>
<td></td>
<td>EG-2 (AR + scaffoldings)</td>
<td>14</td>
<td>8.07 (1.43)</td>
<td>0.38</td>
<td></td>
</tr>
<tr>
<td>Mental load</td>
<td>CG</td>
<td>14</td>
<td>3.71 (1.72)</td>
<td>0.46</td>
<td>0.776</td>
</tr>
<tr>
<td></td>
<td>EG-1 (AR)</td>
<td>14</td>
<td>3.36 (2.49)</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td></td>
<td>EG-2 (AR + scaffoldings)</td>
<td>14</td>
<td>2.79 (1.62)</td>
<td>0.43</td>
<td></td>
</tr>
</tbody>
</table>

4.3 Learning Experience

ANOVA was utilized to compare the difference of learning experiences among the three groups. The results are shown in Table 4. There was significant difference between three groups in learning experience ($F(2, 38) = 4.688, p = .015$). Specifically, the experimental group’s learning experience was significantly higher than the control group. However, no significant difference was found between the two experimental groups. In addition, it is worth noting that students’ learning experience scores in the experimental group reached almost full score (7 points), indicating that the students had a good learning experience in the process of learning with AR application.

### Table 4. The ANOVA Results of the Students’ Learning Experience

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>M (SD)</th>
<th>SE</th>
<th>F</th>
<th>Post-hoc</th>
</tr>
</thead>
<tbody>
<tr>
<td>CG</td>
<td>14</td>
<td>5.32 (.61)</td>
<td>0.16</td>
<td>4.688*</td>
<td>EG1 &gt; CG*, EG2 &gt; CG**, EG2 &gt; EG1</td>
</tr>
<tr>
<td>EG-1 (AR)</td>
<td>14</td>
<td>6.13 (1.07)</td>
<td>0.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EG-2 (AR + scaffoldings)</td>
<td>14</td>
<td>6.22 (0.82)</td>
<td>0.22</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* $p < .05$, ** $p < .01$, *** $p < .001$

4.4 Interview Results

We used open-ended coding methods to code the interview data. The interview results are as follows.

4.4.1 AR Learning vs. Traditional learning

As shown in Table 5, when asked about the perception between learning with AR and traditional methods, students’ answers revealed advantages of AR learning over traditional learning in the following four aspects: (1) cognitive, (2) ability, (3) affective, (4) usability.

For cognitive benefits, AR could promote the retention of scientific concepts, understanding and reflection of scientific concepts. Because AR can make objects stereoscopic and intuitive, which help students understand the concepts more clearly. As for ability, the AR application could involve students in the experiment, so it could improve their hands-on and observation skills, and develop their imagination to some extent. In addition, some students can connect the learning content with their real lives, such as S01, indicating that students could combine previous life experience and have a deeper construction of learning content. And for affective benefits, AR stimulated students’ learning interest in science, improved learning motivation and self-efficacy. Last, for usability, because AR application is small and convenient, students would have more opportunity to do the science experiment. Some students may not have chance to do experiment by themselves in traditional science classrooms.

### Table 5. Examples of Students’ Opinions Related to the Benefits of AR Application

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>M (SD)</th>
<th>SE</th>
<th>F</th>
<th>Post-hoc</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

365
4.4.2 AR with Cognitive Scaffoldings vs. AR without Scaffoldings

When asked about the cognitive scaffoldings embedded in the AR application, students’ answers showed different attitudes towards the cognitive scaffoldings (Table 6). On the one hand, some comments revealed that the cognitive scaffoldings embedded in AR have positive cognitive and affective effects. On the other hand, some students said that cognitive scaffoldings were not necessary, and they don’t need them in learning activities. This may explain why cognitive scaffoldings had no significant effect on students’ learning performance, learning experience, and cognitive load in the quantitative results.

Table 6. Examples of Students’ Different Attitudes Towards Cognitive Scaffoldings

<table>
<thead>
<tr>
<th>Number</th>
<th>Examples of students’ opinions</th>
<th>Coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>S04</td>
<td>It (scaffolding) helps me understand better.</td>
<td>Cognition</td>
</tr>
<tr>
<td>S03</td>
<td>When I forget the key information, I can look at the &quot;Learning tips&quot; button, which helps me a lot.</td>
<td>Positive</td>
</tr>
<tr>
<td>S05</td>
<td>When I got it right, the box (feedback) gave me a sense of accomplishment.</td>
<td>Affective</td>
</tr>
<tr>
<td>S02</td>
<td>I think it’s good and interesting, I can get a few more cues.</td>
<td>Not necessary</td>
</tr>
<tr>
<td>S01</td>
<td>I don’t need a prompt. The prompt is generally to tell you something, I don’t think I need it.</td>
<td>Negative</td>
</tr>
</tbody>
</table>

4.4.3 Difficulties When Learning with AR

Apart from the advantages mentioned above, there are also some difficulties including usability and cognitive problems when learning with AR. Table 7 shows the difficulties when using AR. For usability, marker problems and health problems are two common challenges. And for cognitive
difficulties, the teacher highlighted that AR might be too immersive for students to concentrated on teachers speaking.

Table 7. Examples of Students’ and Teacher’s Responses about Difficulties When Using AR

<table>
<thead>
<tr>
<th>Number</th>
<th>Examples of students’ opinions</th>
<th>Coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>S01</td>
<td>The three colors of light are not very stable.</td>
<td>Marker problems</td>
</tr>
<tr>
<td>S03</td>
<td>The recognition picture is too small.</td>
<td>Usability</td>
</tr>
<tr>
<td>S04</td>
<td>I have problems when I recognize the card.</td>
<td></td>
</tr>
<tr>
<td>T01</td>
<td>It may not good for students’ eyes when use it for a long time.</td>
<td>Health problems</td>
</tr>
<tr>
<td>T01</td>
<td>The following situations often occur during the class: Some students have been addicted to AR’s exploring activities, and they cannot hear me when I told them to do other things.</td>
<td>Distract attention Cognition</td>
</tr>
</tbody>
</table>

5. Discussion and Conclusion

5.1 AR Can Improve Students’ Scientific Achievements and Learning Experience Significantly.

One of the purposes of this study is to investigate whether the AR applications have positive effects on students’ scientific achievements and learning experience compared with traditional teaching methods. The results of this study showed that using AR applications for teaching had significant positive effects on improving students’ scientific knowledge achievements and learning experience. For scientific achievements, our results were consistent with the findings of Chang and Sahin. (Chang, Hsu, Wu, & Tsai, 2018; Sahin & Yilmaz, 2020). This may be caused by AR’s special features, which is that AR can make the objects or phenomena that are difficult to visualize three-dimensional. Therefore, it was easier for students to understand complex and abstract concepts. In addition, learners could directly interact with the content and control the physical objects by themselves, which is more visualized than traditional method. This could explain why AR can improve students’ scientific achievements and retain scientific knowledge. Further, we can also see words like "clear" and "intuitive" in the qualitative data, which verified this conclusion.

As to learning experience, this study indicated that teaching with AR applications can effectively improve students learning experience, which was consistent with the results of Díaz’s research (Díaz, Hincapié, & Moreno, 2015). AR provided the opportunity to interact with virtual objects, bringing a different learning experience, which created a sense of realism for learners. Through the interview data, we can see that students’ favorite part was the inquiry activities in AR. At this part, students had a high degree of freedom and had a lot of opportunities to interact with AR. The students’ responses like "challenging", "I like interaction " indicated the positive experience that AR brought to them. Therefore, the inquiry activities in AR may be a main reason for improving students’ learning experience.

5.2 Cognitive Scaffoldings Embedded in AR Have No Significant Impact.

The second purpose of this study is to examine the effect of the cognitive scaffoldings embedded in AR applications on students’ scientific achievements, learning experience, and cognitive load. The results showed that there was no significant difference between the two experimental groups in all dependent variables. Both EG-1 and EG-2 had high scientific achievements, learning experience and a low cognitive load. This was not consistent with Wu’s research results (Wu et al., 2017). Wu developed an AR learning system based on mindtool to help students build their knowledge in science classes, showing that this kind of scaffolding can effectively improve students’ learning performance. In his study, a low cognitive load was also found in both control group and experimental group during learning activities, and there was no significant difference between the two groups, which is consistent with our study.

This can be explained by the following reasons: First, some of the cognitive scaffoldings in this study may not be necessary for students to promote their learning. The qualitative data showed that
there were differences attitudes towards cognitive scaffoldings embedded in AR, which was consistent with Yoon’s research results (Yoon et al., 2012). Yoon explored the effect of integrating cognitive scaffolds to AR system, indicating that some cognitive scaffolds may not have been necessary to increase learning of general concepts except for digital augmentation, which was consistent with the qualitative results of this study. Another reason can be fact that the cognitive scaffolds provided in this study might not be well connected to the difficulties faced by students. In multimedia learning, cognitive and emotional design standards should be followed when developing an educational application, and giving special feedback and suggestions according to student problems are very necessary too. Therefore, it is worth discussing about how to design the scaffolds in AR in detail. In the future, researchers can explore the impact of the way in which scaffolds embedded when teaching with AR. For example, what’s the effects of different kinds of scaffolds and the ways we used them. This would be helpful for us to know how to design the scaffolds in AR to help students learning in the future.

Acknowledgements
Our work is supported by the National Natural Science Foundation of China (61977007) and 2021 Comprehensive Discipline Construction Fund of Faculty of Education, Beijing Normal University.

References


Development and Field Testing of a MALL for Filipino with a Reusable Framework for Mobile-Based Drills


*Ateneo Laboratory for the Learning Sciences, Ateneo de Manila University, Philippines
*bDepartment of Education, Ateneo de Manila University, Philippines

*jagapito@ateneo.edu

Abstract: This paper describes the development and field testing of *Ibigkas!* Filipino, a mobile game that exercises learners’ fluency in identifying synonyms (*kasingkahulugan*) and antonyms (*kasalungat*) in the Filipino language. Twenty-four students from Grades 4, 5, and 6 were invited to play and answer comprehension tests to determine whether the game helped them improve their understanding of the content. Self-report questionnaires assessed the extent to which they enjoyed it. Additionally, three teachers were invited to a focus group discussion (FGD) to gather their insights about the game and how they may use it in their classes. Self-report feedback from students showed the game was fun, interesting, and sufficiently challenging. A significant increase in the post-test comprehension scores of the Grade 4 participants was found. This shows the potential of the game to make learning fun while helping realize learning goals. Teachers indicated they can use the game to supplement their Filipino classes and that the students will be receptive to the idea of utilizing a game for learning.

Keywords: game-based learning, Filipino comprehension, Philippines, mobile-assisted language learning

1. Context and Motivation

Filipino is the official and national language of the Philippines (The 1987 Constitution of the Republic of the Philippines, Article XIV). It is based on Tagalog, one of the country’s major languages. The 1987 Philippine Constitution stipulates that the Philippine government is to “take steps to initiate and sustain the use of Filipino as a medium of official communication and as a language of instruction in the educational system” (The 1987 Constitution of the Republic of the Philippines, Article XIV). There are policies that have been established to fulfill this end (DO 81, S. 1987). At present, the Basic Education Program are guided by the Policy Guidelines on the K to 12 Basic Education Program (DO 21, S. 2019). This order states that the medium of teaching and learning from Grades 1-12 are both Filipino and English. Also, Filipino subjects are part of the core learning areas in all grade levels.

In spite of these efforts, Filipino learners still have difficulty in speaking, reading, and writing in Filipino. Many learners who are born and raised in the Philippines use English as their first language (Roque et al., 2017). Hence, many of these learners have difficulty speaking, reading, and writing in Filipino. When in school, they learn Filipino as a second language, often needing English translations of texts to increase comprehension (Roque et al., 2017). These deficiencies are also evident in the results of within-country tests. The Philippines’ Department of Education (DepEd) Regional Office 02 (Cagayan Valley) reported that their 2018 NAT results for Grade 6 show that the mean performance by subject area were far below the 75% acceptable MPS with scores for Filipino averaging 51.68% (DepED RO2, 2019). Additionally, Proficiency Level Distribution in the subject shows that about 95% of the examinees did not reach the proficiency level for Filipino with only 0.02% and 5.65% being highly proficient and proficient, respectively (DepED RO2, 2019). Apart from the wide influence of Western culture, other factors may contribute to. In a PETS-QSDS (Public Expenditure Tracking Surveys-Quantitative Service Delivery Surveys) study conducted in 2014 involving a nationally
representative sample of 946 Grade 10 high school teachers, roughly 38% mentioned the need for better facilities and about 60% expressed the need for additional teaching materials (Al-Samarrai, 2016).

As a response to the need for additional learning materials, the Ateneo Laboratory for the Learning Sciences has developed mobile phone-based games that can be used to support the learning process in Philippine classrooms. In this paper, we describe the development and field testing of *Ibigkas! Filipino*, a mobile game that exercises learners’ fluency in identifying synonyms (*kasingkahulugan*) and antonyms (*kasalungat*) in the Filipino language. The purpose is two-fold: (1) to describe *Ibigkas! Filipino* and how it was built using a framework for creating mobile collaborative, drill-and-practice type games; and (2) to assess the target audience’s response to the game.

2. Mobile Learning

Mobile learning is becoming popular in educational environments because of its mobility and portability. Through the use of mobile devices, learning can take place anytime and anywhere, breaking past the temporal and spatial limitations of traditional learning methods (Christensen & Knezek, 2017). Developing countries have recorded mobile phones to be their primary computing platform. As of 2018, the developing world had an estimated 102.8 mobile phones for every 100 people (ITU, 2018). Consequently, teachers and students continue to explore integrating this technology in the teaching and learning process. This is apparent in the Philippines. In a survey of 710 Grades 4, 5, and 6 students from two elementary state schools in Metro Manila, it was found that they have access to cellular phones (63%), computers (54%), and tablets (36%) (Rodrigo et al., 2019b). This shows the mobile platform is a promising venue for delivering learning support to a wider range of users. Hence, *Ibigkas! Filipino* was implemented for use on mobile devices such as cell phones and tablets.

3. Digital Game-Based Learning

Digital Game-Based Learning (DGBL) is a learning approach that incorporates educational content into digital games (Chang & Wang, 2019). It is a method that utilizes the entertainment capabilities of games in learning environments. The growing popularity of using games as tools in education may be attributed to the motivational affordances (e.g. immediate feedback, quests) inherent to them. They can create fun experiences that stimulate positive attitudes and emotions among learners while assisting in the realization of learning goals. DGBL is found to be common in teaching knowledge and skills in different topics such as science, mathematics, and language learning, among others.

3.1 Collaborative Learning Games

Collaborative learning (CL) is defined as “a fundamentally social process of knowledge-building”, in which learners are provided an environment where they can work together to complete a task while communicating with one another during the process (Kukulska-Hulme & Viberg, 2018). CL can take place between two learners, or even a learner and a teacher (Kukulska-Hulme & Viberg, 2018). CL necessitates learners to articulate their thoughts to contribute to the groups’ mutual pursuit of a learning task (Koschmann et al., 1996). Collaborative games are activities in which players who share a common goal work together in order to share payoffs and outcomes (Zagal et al., 2006). The consequences of an individual decision, whether positive or negative, are shared by all the members of the group. Research has shown that collaborative games are fun, engaging, and motivating (Wendel et al., 2013), hence, able to promote positive learning attitudes, achievement, and self-efficacy.

3.2 Mobile-Assisted Language Learning (MALL)

The emergence of mobile learning has reached many different subjects, including that of language learning. Mobile-Assisted Language Learning (MALL) is an area within mobile learning that focuses on language learning topics such as grammar, listening and reading comprehension, and vocabulary acquisition (Sung et al., 2015). In a systematic review conducted by Pesson and Nouri (2018), it was found that most research in MALL focus on improving L2 (Second Language) proficiency. Additionally, their review showed how MALL was commonly used in formal education, typically
included as a support to existing curricula. This warrants other researchers’ perspective that MALL applications can be used to support learning in both formal and informal contexts (Sung et al., 2015).

4. Ibigkas! Filipino

Ibigkas! Filipino is a game for practicing identification synonyms (*kasingkahulugan*) and antonyms (*kasalungat*) in the Filipino language. It was built using a framework for creating mobile collaborative, drill-and-practice type games for both Android and iOS devices. The framework is flexible enough to accommodate any subject area in which learners have to identify what items go together or map to each other. Other games that were developed to test the framework’s flexibility are Ibigkas! – a game that helps learners develop fluency in identifying rhymes, synonyms, and antonyms in English (Rodrigo et al., 2019b) and Ibigkas! Math – a game for learning arithmetic (Rodrigo et al., 2019a).

Ibigkas! Filipino was created as a response to teachers’ feedback from previous experiments to create a game for teaching the Filipino language. Like the other Ibigkas! games, it has two modes: single-player and multi-player. In the single-player mode, the game starts by asking the player to choose a content mode (*kasingkahulugan* or *kasalungat*) and a level of difficulty. When the game begins, the player receives a target word and a list of three word choices. The player must select the word that is the best match to the target word. The multiplayer mode allows groups of up to 8 to play together. Each player must have a mobile phone that is connected to a network (not necessarily the Internet). At the start, the host player selects the content mode (*kasingkahulugan* or *kasalungat*) that they will play. When a game round begins, a random player from the team receives the target word (Figure 1a). All players will then receive lists of words, only one of which is the correct answer that corresponds to the target word (Figure 1b). The player who has the target word has to say it aloud (*Ibigkas* is Filipino for ‘speak up’) so the other players can know what it is. All other players then check their list of words to see if they have the correct answer. The player with the correct answer should likewise say it aloud and tap it on his screen. Once the correct answer has been tapped, the round ends and a new one begins.

![Figure 1](image.png)

5. Testing Methods

The field test for Ibigkas! Filipino was conducted at a public school in Metro Manila, Philippines. Students from Grades 4, 5, and 6 were invited and grouped by grade level during the study. Separate testing sessions for each grade level were ran due to the limited number of available mobile phones. The participants were given an orientation about the study and the details of their participation. Then, they answered a demographics questionnaire that determined their level of access to mobile phones. It also tried to assess their usage, attitude, and perception towards the Filipino language by giving them 8 statements to which they would specify their level of agreement (1 = Strongly Disagree to 5 = Strongly Agree). Some sample statements are “I speak Filipino at home.” and “Learning Filipino is important.” After completing the demographics questionnaire, students answered a 12-item multiple choice pre-test that tried to assess their ability to determine the correct synonym (*kasingkahulugan*) or antonym (*kasalungat*) of a given Filipino word. Once done, they were asked to play Ibigkas! Filipino for about 20 minutes. After which, a 12-item multiple post-test was given to evaluate whether they were able to improve their Filipino language comprehension skills through the game. The comprehension tests were based on the pool of questions in the game. Lastly, they were asked to complete a two-part questionnaire with items adapted from the GBL Engagement Metric (Chew, 2017) and the IMI questionnaire (Ryan, 1982).
6. Analysis

6.1 Participant Profile

A total of 24 students participated: 10 from Grade 4 and 7 each from Grades 5 and 6. Majority of the Grades 4 and 6 participants owned cell phones. However, only 2 of the 7 Grade 5 students had their own. All of them played cell phone games and most were into playing educational games. See Table 2.

Table 2
Profile of Participants

<table>
<thead>
<tr>
<th></th>
<th>Grade 4</th>
<th>Grade 5</th>
<th>Grade 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>M 5</td>
<td>F 5</td>
<td>M 3</td>
</tr>
<tr>
<td>Age Range</td>
<td>9 - 11</td>
<td>10 - 11</td>
<td>11 - 12</td>
</tr>
<tr>
<td>Had their own cell phones</td>
<td>8 (80%)</td>
<td>2 (26%)</td>
<td>5 (71%)</td>
</tr>
<tr>
<td>Played games on cell phones</td>
<td>10 (100%)</td>
<td>7 (100%)</td>
<td>7 (100%)</td>
</tr>
<tr>
<td>Played educational games on cell phones</td>
<td>6 (60%)</td>
<td>7 (100%)</td>
<td>7 (100%)</td>
</tr>
</tbody>
</table>

When asked about their usage, attitudes, and perception towards the Filipino language, students said they spoke Filipino at home (4.5/5.0) or with their friends (4.6/5.0). They did enjoy learning (4.6/5.0) and reading (4.6/5.0) in Filipino. They expressed a desire to learn Filipino (4.4/5.0) and agreed that it is important (4.6/5.0). These show that the students tend to use Filipino as a medium of communication and are inclined to improve their skills and knowledge of the language.

6.2 Game Engagement Results

Ibigkas! Filipino generally stimulated positive feelings in the students as per their self-report responses. Results of the GBL questionnaire show that they followed instructions carefully (4.5/5.0) and tried their best to identify the correct synonym (kasingkahalongan) or antonym (kasalungat) of the target word (4.7/5.0). The game allowed them to utilize their previous knowledge (4.8/5.0) and further improve their Filipino vocabulary (4.6/5.0). Playing the game interested them (4.6/5.0) and that they look forward to finishing each level (4.7/5.0). The Grade 4 students seemed to be more open to asking questions when they were lost (4.3/5.0) than the Grade 5 (3.57/5.0) and Grade 6 (3.71/5.0) students. Generally, they said the game is fairly challenging (3.7/5.0). The IMI results showed the students enjoyed the game a lot (6.96/7.0). Students said it was fun to play (6.96/7.0), with scarcely any who thought it was boring (1.9/7.0). It was important for them to do well in the game (6.6/7.0) and so they tried their best to answer correctly (6.8/7.0). They expressed uncertainty with regards to how skilled they were in the game (4.8/7.0). Nonetheless, they were satisfied with how they performed (6.8/7.0).

6.3 Comprehension Scores

Part of the study was to examine whether playing the game improved the students’ performance. Grade 4 students’ pre-test scores averaged 59%, Grade 5 students averaged 60%, and the Grade 6 students averaged 68%. While it was somehow expected that the higher grade levels have higher average scores, a single-factor ANOVA showed that the differences among the groups were not significant (F(2,21)=0.56, p=0.58). The post-test scores show an increase with the Grade 4 students averaging at 77%, Grade 5 at 71%, and Grade 6 students at 75%. Paired sample t-tests to compare the pre- and post-test scores per grade level were performed to determine if the differences were significant. Table 3 shows that the scores of the Grade 4 participants before (M=7.1, SD=2.77) and after (M=9.2, SD=2.84) playing Ibigkas! Filipino was significant (t(18)=−2.8, p=0.006). However, no significant differences were found between the pre- and post-test scores of the Grade 5 and Grade 6 participants.
Table 3

Paired Sample T-Test Results Per Grade Level

<table>
<thead>
<tr>
<th>Grade</th>
<th>N</th>
<th>Pre-Test mean</th>
<th>Pre-Test stdev</th>
<th>Post-Test mean</th>
<th>Post-Test stdev</th>
<th>df</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 4</td>
<td>10</td>
<td>7.1</td>
<td>2.77</td>
<td>9.2</td>
<td>2.84</td>
<td>18</td>
<td>-2.8</td>
<td>0.006</td>
</tr>
<tr>
<td>Grade 5</td>
<td>7</td>
<td>7.14</td>
<td>4.48</td>
<td>8.57</td>
<td>1.62</td>
<td>10</td>
<td>-1.53</td>
<td>0.08</td>
</tr>
<tr>
<td>Grade 6</td>
<td>7</td>
<td>8.14</td>
<td>7.81</td>
<td>9</td>
<td>2</td>
<td>9</td>
<td>-0.72</td>
<td>0.24</td>
</tr>
</tbody>
</table>

These results suggest that the game may have helped improve the performance of the younger participants in the tests given. Given the small N, these findings have to be interpreted with some caution. They imply that that the game has the potential to support learning of synonyms and antonyms in Filipino. However, more testing has to be performed for the results to be conclusive.

6.4 Focus Group Discussion (FGD) with Teachers

Three (3) teachers (one representative per grade level) were invited to participate in a focus group discussion (FGD) to determine the group activities they give their students that are focused on Filipino learning. The conversation likewise sought to collect their insights regarding Ibigkas! Filipino. Group activities usually come in the form of (1) role-plays; (2) composition (as in poems); poster-making; and (4) sentence construction based on pictures. The students’ capabilities and the difficulty of the activities often determine how they are grouped. Teachers would sometimes give incentives to encourage participation. Students who do not contribute to group work are often given grade deductions, assigned reporting tasks, or asked to do a separate activity. The teachers generally liked Ibigkas! Filipino. They acknowledged how learners would enjoy playing it. They did not specify anything they dislike about the game, but they had suggestions such as setting a fixed time for every difficulty level and expanding the content (e.g. Filipino pronouns). They are interested in using it in their classes and they agreed it can be among the group activities they facilitate. They also think their students will enjoy the multi-player mode because it can be organized such that multiple groups compete to get the highest score. The added competitive element together with the collaborative nature of the game makes it more exciting.

7. Conclusion

This paper presents the implementation and testing of Ibigkas! Filipino, a collaborative, mobile phone-based, drill-and-practice game that helps learners develop fluency in identifying synonyms (kasingkahulugan) and antonyms (kasalungat) in the Filipino language. The students’ self-report responses indicate positive feelings towards the game. They found it interesting and fun to play. They exerted effort in following instructions and identifying the correct match to the target words. It was important that they performed well, hence, they tried their best to answer correctly. The students expressed interest in improving their knowledge of the Filipino language. The comprehension scores turned out to be generally good. A significant improvement in the post-test scores of the Grade 4 students has been found. This indicates how the game may have helped improve the scores of the younger students. Their openness to ask questions whenever they felt lost might have contributed to this since they were somehow able to obtain feedback. Though not significant, an increase in the mean post-test scores of the Grades 5 and 6 participants was also noted. However, these results should be interpreted with caution due to a small N. Further testing is required to arrive at more conclusive results. Still, this study shows the potential of the game as an aid for learning Filipino synonyms and antonyms.

This work is an attempt to address the need for additional learning materials in the classroom. The feedback from the students demonstrate how the game can contribute to making the learning process more fun and interesting. The improvement in post-test scores, particularly for the Grade 4 students, is encouraging. Teachers’ suggestion to expand the content is an indication that they see the value of these kinds of materials. As previously mentioned, the mobile platform is the most accessible to a wider range of audience, including those who may be underprivileged. The development of the Ibigkas! suite of games was driven by this so that more teachers and students will be encouraged to use them to make the realization of learning goals more fun and exciting.
Acknowledgements

We thank the Ateneo de Manila University; the principals, teachers, volunteers, and learners of our partner public schools for their participation; Ma. Rosario Madjos; and the Commission on Higher Education and the British Council for the grant entitled JOLLY: Jokes Online to improve Literacy and Learning digital skills amongst Young people from disadvantaged backgrounds.

References


Enhancing Customers’ Knowledge and Decision Making using Augmented Reality

Wisanoo BOONRAT\textsuperscript{a}, Vimita VAIDYA\textsuperscript{a}, Nilufar BAGHAEI\textsuperscript{b}\textsuperscript{*}, Hamid SHARIFZADEH\textsuperscript{c}, Ali AHMED\textsuperscript{d}
John CASEY\textsuperscript{c}

\textsuperscript{a}Department of Information Technology, Otago Polytechnic Auckland Campus, New Zealand
\textsuperscript{b}School of Natural and Computational Sciences, Massey University Auckland, New Zealand
\textsuperscript{c}School of Computing, Unitec Institute of Technology, New Zealand
\textsuperscript{d}School of Engineering and Computer Science, Victoria University of Wellington, New Zealand
\textsuperscript{*}N.Baghaei@massey.ac.nz

Abstract: Augmented reality has seen massive success in recent years as it provides an opportunity for a seamless and rich user interaction with the real world. Recent studies have shown augmented and virtual reality can play a significant role in the future of retail. Innovations can help customers make good choices and can improve their confidence and satisfaction in their purchasing decisions. It also reduces the physical interaction between customers and store especially during lockdowns (i.e. during pandemics such as COVID-19). This paper presents ARiel, an augmented reality application for mobile commerce (m-commerce), enabling customers to turn catalogued items into 3D animated models. It was developed using Google’s ARCore platform. Customers can try on products such as jewellery or watches using their mobile phones without being physically present in the retail shop. Data about their purchasing decisions are logged in a database. We describe the process of designing and implementing ARiel and the outcome of an evaluation using Nielsen’s ten usability heuristics. We also outline our plan for future work.

Keywords: Augmented reality, m-commerce, usability heuristics

1. INTRODUCTION

E-commerce market will see a shift in purchasing power through the next decade due to internet penetration, especially on mobile devices (Statista, 2019). There are many websites such as TradeMe, Amazon, eBay and Alibaba providing the various ranges of products with details, images, and reviews to help customers with their purchasing decisions. Reviewing the product’s images and consumer reviews in blogs, twitter and wikis are common (Tirunillai & Tellis, 2012). The customers can, therefore, presume the feeling of product possession before making a purchase. Due to the extensive use of smartphones, consumer online-posting has become a vital factor in product marketing (Chen et al., 2011; Singh et al., 2018). The electronic word of mouth, rating, reviews, forums and communities can shape consumer trust, which is essential for a business to successfully promote the products (Hajli et al., 2014). However, fake reviews can be formulated to gain competitive advantage by the competitors (Singh et al., 2018). Despite many product details, images and consumer reviews provided on the Internet, we believe there is still a need for the customers to try the products by themselves.

There has been a significant interest in Virtual Reality (VR) and Augmented Reality (AR) technologies in recent years. Both VR and AR can be utilized in many application domains such as health care, business, education, and amusement (Alkhamisi & Monowar, 2013). AR can help the user experience (UX) in three fundamental ways (Li & Fessenden, 2016):

1) By decreasing the interaction cost to perform a task:
The user can remain in the current environment and have relevant data displayed there, without doing any special action. In contrast, with a non-AR user interface (UI), the user needs to take an explicit action to access the information, which could require extra effort on the user’s part. The lack of commands in AR interfaces makes the interaction efficient and requires little user effort.
2) By reducing the user’s cognitive load:
In the absence of an AR system, the user would have to remember not only how to use the smartphone or the desktop to find the information they need, but also go through the process of finding it. With an AR system, the useful part of the information is displayed automatically, and the user does not need to add extra information to working memory or spend the effort to save it on paper or elsewhere.

3) By combining multiple sources of information and minimising attention switches:
With AR systems, the relevant information is displayed in an overlay on top of the object itself, so the user will not need to divide their attention. Many complex tasks (e.g., surgery, writing a report) do involve putting together multiple sources of information; some of them will benefit from AR.

There has been some research on VR and AR shopping assistant applications, such as Virtual Fitting Rooms (VFRs) which uses Natural Interaction (NI) technology with Microsoft Kinect and Asus Xtion as an environment for NI applications (Pachoulakis & Kostas, 2012). Pachoulakis and Kostas concluded that the VFRs allowed the customers to try-on apparel and mix-and-match accessories without being physically present in the retail shop and reduced much of the guesswork involved in shopping, which resulted in an enhanced shopping experience. Although this approach is interesting, most consumers could not try the VFRs platforms, because it required Microsoft Kinect or Asus Xtion which, people may not generally have access to. Waterlander et al. (2015) researched and implemented the Virtual Supermarket. People could experience and interact with the VR supermarket to purchase virtual ford products. Waterlander et al. claimed that shopping patterns in the Virtual Supermarket were comparable to shopping in real life. Other development using AR has been applied on the mobile application to support shoppers to find healthy food products in a grocery store aisle, reducing the amount of time to find desired healthy food products, e.g. (Ahn et al., 2015).

Figure 1. Screenshots of Ariel's product information pages

Research on the global online retail market indicates that the market value has demonstrated consistent growth for many years, and it is expected to reach $26 trillion in 2020 and $29 trillion in 2023 respectively (eMarketer, 2019). Despite VR and AR technologies gaining a lot of traction among consumers and businesses, these immersive technologies are growing at different rates (eMarketer, 2019). It is anticipated that 51.8 million people in the US will use VR and 77.7 million will use AR at least once per month in 2020.

The introduction of Apple’s ARKit and Google’s ARCore in 2017 signalled the tech industry’s confidence in the ongoing support of AR experiences (eMarketer, 2019). Google has introduced the ARCore platform for building augmented reality experience, which enables android phones to sense the AR environment (Google, 2019). With ARCore, the android phones can detect the real-world environment and place virtual objects on a flat surface or augmented image (i.e. a specific 2D image). Recent studies have shown the effectiveness of this platform in increasing user engagement with an interface. For example, Alvaro-Tordesillas, Crespo-Aller, and Barba (2019) used ARCore to design and develop a mobile application (ArtAlive) for the generation of augmented reality
experiences on museum sculptural objects. They reported that after the use of ArtAlive, the level of involvement with the exhibitions exceeded the typical audio-guide or printed brochure. Zhang, Yao, Zhu and Hu (2019) proposed an assistive navigation system for visually impaired people that took advantage of ARCore, path planning, and human-machine interaction to provide fluent and continuous guidance in avoiding obstacles and risky places.

Current research shows that there are opportunities to develop e-commerce applications in mobile devices (m-commerce), leveraging the AR technology, to improve the online shopping experience. However, there is little work done on evaluating the usability of AR interfaces. We designed and developed a prototype for an m-commerce application that allows users to select and view the products’ details similar to other online shopping applications. We then provided the users with an option to try-on products to visualise how the products would look like using a 3-D model. The research question we investigate in this paper is whether our augmented reality m-commerce application can meet Nielsen’s top ten usability heuristics.

2. Designing ARiel: An AR Jewellery M-Commerce Application

ARiel is a m-commerce application with augmented reality “try-on” feature, which aims to enhance the users’ shopping experience by allowing them to try the 3-D products using their mobile phone. It has two main components: e-commerce android application and marker-based AR feature. ARiel’s UI design was based on Jakob Nielsen’s ten usability heuristics principles (Nielsen, 1994), described further in Section 3.

2.1 M-commerce Android Application
The e-commerce Android application part of ARiel consists of three pages, which are product category, product list, and product detail page. The users have to select the desired product before trying it with AR try-on feature. Its look-and-feel is similar to general applications in the Android market.

2.2 Marker-based AR Application
The AR application part of ARiel is a marker-based AR, which anchors AR contents (3-D objects) to a specific visual marker. There are two pages here, the guide page outlining the steps on how to print the marker and use AR camera, and the AR camera page. The AR feature requires the user to print a marker and attach the marker to fingers or wrist. They will then have to see the marker through the mobile phone’s camera. The 3-D object of the products will be placed on top of the marker. The application also provides other products for the users to try on via this screen (see Figure 2).

We chose ARCore as a software development kit for the AR try-on feature, as one can simultaneously implement the AR try-on feature and the e-commerce mobile application without dependencies between each other. We, therefore, demonstrated that any current android applications in the market have the feasibility to use AR technology in their applications.

3. Usability Evaluation

Jakob Nielsen proposed 10 general principles for interaction design (Nielsen, 1994), that have been extensively used for designing user interfaces over the last 25 years. They are referred to as “heuristics”, as they are “broad rules of thumb and not specific usability guidelines”. We used these usability heuristics to evaluate ARiel’s UI:

1. **Visibility of system status**: According to the first heuristic, the system should always keep users informed about what is going on, through appropriate feedback within a reasonable time.
   ARiel users are informed about the current application’s states.

2. **Match between system and the real world**: The system should speak the users’ language, with words, phrases and concepts familiar to the user.
   All icons and messages in ARiel are meaningful and concise. The users can find them in everyday life, such as the camera icon and labels on the page’s titles.
3. **User control and freedom**: Users often choose system functions by mistake and will need a clearly marked “emergency exit” to quickly leave the unwanted state. The back and home buttons are provided for emergency exit.

4. **Consistency and standards**: Users should not have to wonder if different words, situations, or actions mean the same thing. The position of buttons of all activities in the application are consistent and follow the same pattern of other commonly used products in the Android market.

5. **Error prevention**: Careful design can prevent a problem from occurring in the first place. Either eliminate error-prone conditions (ideally) or check for them and present users with a confirmation option before they commit to the action. ARiel would aim to stop users from making errors and warning messages are displayed at the appropriate times.

6. **Recognition rather than recall**: Minimise the user’s cognitive load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. The selected product will be highlighted in the app when it is displayed on the product list of AR try-on screen.

7. **Flexibility and efficiency of use**: Accelerators may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. The system guide will be provided to the users who have no experience of using AR camera, which can be easily be skipped. Shortcuts are also provided for expert users.

8. **Aesthetic and minimalist design**: Dialogues should not contain irrelevant information. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility. The use of colour, shape, and typography in the application has been developed under the concept of material design aimed to provide minimum information displayed in each screen throughout the application.

---

*Figure 2. Screenshots of ARiel's product information page*

---

9. **Help users recognise, diagnose, and recover from errors**: Error messages should be expressed in plain language with no source code, precisely indicate the problem, and constructively suggest a solution for the user.
ARiel’s error and warning messages are easy to understand for both novice and expert users. It shows users how to fix the problem, so they can carry out interacting with the interface.

10. **Help and documentation:** Although it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user’s task, displayed when necessary, list concrete steps, and not be too large. ARiel provides a single-page user guide, written in simple plain language, to educate the users on how to use AR try-on feature.

The usual approach to designing a mobile application that makes use of the phone camera is to turn on the camera at the beginning when the application starts. The users will then operate the application via the given menu on the camera screen. The main disadvantage of this approach is that mobile devices have limited screen size, making it difficult for e-commerce applications to provide all product information along with menus in a single camera screen. Our proposed approach is more practical for the existing online e-commerce applications in the Android market, as the AR try-on module can be developed separately, and then be integrated into the existing application. The retailers can slowly introduce and educate the current users about this new feature.

4. **CONCLUSIONS AND FUTURE WORK**

In this project, we designed and developed an AR try-on feature to enhance the consumers’ shopping experience. Due to COVID-19 restrictions and the major challenges around user studies, we were not able to conduct an in-person evaluation to study participants’ user experience and whether our app enhances their knowledge and decision making. We, therefore, used a well-known usability framework to validate the design of the interface.

AR interfaces provide great opportunities for improving user experience. The future of retail is profoundly informed by understanding what is different and what is similar in real and virtual worlds and how immersive technologies can affect both. Innovations can help customers make good choices, experience fewer time constraints or even increase confidence and fulfilment in their decisions. Retailers, however, can gain the advantage of emerging technologies to increase the engagement of their customers and simplify their lives. Our future plan is to conduct an evaluation study to examine the effectiveness of ARiel on improving customers’ shopping experience and knowledge and the number of purchasing decisions made as a result of using the AR feature. The information overlaid over the physical world in AR interfaces can initially be overwhelming, but we believe it can help the customers with their purchasing decisions in the long run.

**REFERENCES**


Investigating the effect of using the social semantic tagging-based learning guidance on science learning

Tien-Yu Hsua, Jun-Ming Chena*, Hsin-Yi Lianga
*Department of Operation, Visitor Service, Collection and Information Management, National Museum of Natural Science, Taichung, Taiwan
*jackychen.academic@gmail.com

Abstract: Enhancing prior knowledge has been recognized as an important learning strategy in enhancing the learning experience and performance for students. Past studies on enhancing prior knowledge have focused on reading materials that are manually generated by website administrators and educators. This is time-consuming and expensive process, such that prior knowledge cannot instantly correct and effective way to adaptive aid students in the learning process. To cope with these problems, this paper proposes a social semantic tagging-based learning guidance approach to obtaining relevant prior knowledge and the presented study examines the impact of using the social semantic tagging-based learning guidance on science learning. To evaluate the effectiveness of the proposed approach, an experiment was conducted by assigning 56 students to participate in this learning activity. The students in the experimental group adopted the social semantic tagging-based mobile learning approach, while those in the control group learned with the conventional mobile learning system. The experimental results show that the proposed approach not only promoted the students’ learning achievements and motivations, but also improved their learning self-efficacy and socialization. The proposed approach with prior knowledge construction provides additional help for students and teachers in conducting and participating in interactive English reading learning activities.

Keywords: Ubiquitous learning, social semantic tagging, intelligent tutoring systems, interactive learning environments

1. Introduction

Reading learning can be regarded as a process of accumulating information or experience. Enhancing students’ prior knowledge before reading scientific articles is becoming increasingly important for students, as it helps students learn quickly and effectively. Yang and Quadir (2018) points to a large body of research that indicates “learning proceeds primarily from prior knowledge, and only secondarily from the presented materials”. Therefore, to help educators in enhancing students’ prior knowledge, an effective prior knowledge recommendation and learning guidance tools play an important role to assist students in obtaining the relevant organizational prior knowledge in the learning process (Chung, Hwang, & Lai, 2019; De Medio et al., 2020).

Owing to the recent rapid developments of big data application and social network technologies, new learning technologies are continuously being developed (Shorfuzzaman et al., 2019). Applying social network analysis and semantic analysis enhanced teaching modes are becoming a widely adopted learning mode that enables learning activities to be conducted in learning tasks (Gruzd, Paulin, & Haythornthwaite, 2016; Shen & Ho, 2020). Moreover, educators are increasingly turning to Web 3.0 applications such as semantic web, social networking sites, and semantic wikis to enhance classroom learning. The emergence of Web 3.0 has not only enhanced the social semantic web use of the Internet, but also significantly changed the classroom educational experience (Songkram et al., 2019). This new learning technologies and extend learning strategy support user autonomy through increased levels of socialization and interactivity, access to open communities, and peer-to-peer networking. Previous studies have demonstrated that the social network analysis with prior knowledge construction provides additional opportunities for learning instruction (Gašević et al., 2019; Zambrano et al., 2019).
However, despite the great advantages of social network analysis with prior knowledge construction mentioned above, it is difficult to set up complex knowledge construction and personal learning recommendation to adaptive aid students in the learning process (Holland, 2019; Omodan, 2019). Therefore, it has become an important issue to develop methodologies or tools to assist students in obtaining the relevant organizational prior knowledge effectively. To cope with this problem, this study proposes a social semantic tagging-based learning guidance approach for mobile learning. Moreover, learning motivations, perceived ease of use and usefulness, and students’ self-efficacy are measured to investigate the effects of the proposed approach on the in-field performance of the students from different aspects.

2. A social semantic tagging-based learning guidance approach for supporting prior knowledge recommendation

This study designs a social semantic tagging-based learning guidance approach and develops a prior knowledge recommendation system, called SSTL, which combines with social tagging and semantic concepts of articles to exploit social networking technology to improve a prior knowledge reservoir for students. Figure 1 illustrates the architecture of this approach. The definitions and the formulations of the proposed approach involve various strategies in structural analysis, semantic analysis and social network analysis. The basic ideas of each phrase and constructing process are outlined below.

STEP 1: Extracting the concepts from reading an article by using structural analysis

In order to help students organize ideas through tagging learning, we first considered grasping the structure of the articles and retrieving key information by using the technique of text summarization. Moreover, in the process of summarizing a reading article, the characteristics of summarization can extract the important terms and represent the meaning of the reading article. The information is useful to make students aware that the text has structure, and to provide sufficient practice in the learning process so that they can respond to those clues during the tagging process of our design. After the summarization process is carried out, each concept within the article is used to construct the concept-effect relationships with tag assignments (tag, concept, exam, and user) from individual students.

STEP 2: Using tags to construct relationships of implicit semantic meaning

After all the possible mapped concepts are found for a tag, we need to consider how to find implicit reading clues. In doing so, students’ annotated tags are regarded as a medium for identifying implicit semantic meaning. Thus, we use Latent Semantic Analysis (LSA), a text-based summarization algorithm first proposed by Gong (Garrison & Anderson, 2003), to help construct the implicit semantic meaning of a text. After performing LSA, more semantic relationships can be brought into our tag scoring equation, so that the equation can provide an efficient way to extract the implicit meaning of the students’ tags.

STEP 3: Providing learning guidance by using social network analysis

First, we combine above implicit meaning and topic preferences generated by integrating tag weighting and constructing a semantic similarity matrix between the tags and important concepts within
the reading material to perform spreading activation (SA). Research has demonstrated that the spreading activation approach is employed in many other systems for modeling concepts that might be related (Abbasian & Farokhi, 2019). Moreover, it can help to construct semantic relationships, which are used to draw semantic inferences from a generated vector. After performing SA, the result, therefore, is the generation of all patterns of tags related to a set of article topics, which can represent a view of a learning network in the specific usage context. Meanwhile, it can be seen as a learning mechanism where the system learns from the student’s tagging behaviors. Lastly, in order to discover the appropriate prior knowledge for a student, the final activation vector of the tags examine the prior knowledge database to find the most appropriate tagged article to serve as a prior knowledge supplement. The article with the highest value is then selected as the most appropriate prior knowledge article for the user.

3. A social semantic tagging-based learning environment for knowledge construction

Given the learning guidance approach of our system outlined above, this section also covers the student interface designed to assist students in obtaining the relevant organizational prior knowledge. Figure 2 illustrates the student interface of the social semantic tagging-based learning guidance system, which consists of five areas: (1) the “topics and article content” area located in the left part of the window, which provides students select a learning subject, and then the system interface for reading learning is displayed. Meanwhile, article content is highlighted, and can include the key sentence from the key sentence computation. This highlighted information provides each student with a quick and useful personal snapshot of the reading material; (2) the “tagging” area located in the upper-center part of the window, which provides students utilize the input area to create a list of tags. The interface encourages students to construct meaningful words or phrases to represent the article’s ideas. The students use tags to make a clear overview in their mind for their reading. (3) the “discovering” area located in the lower-center part of the window, which helps students organize, discover interesting clues, and refine their thoughts or ideas of the reading. When students click on a given tag in the tag cloud, the system serves as a useful reference guide, as well as selects suitable prior knowledge materials for students by analyzing the characteristic of the tag cloud. (4) the “discussing whiteboard” area located in the upper-right part of the window for students. Students were given feedback from one of their peers and gives students a chance to sharpen their views on their reading, and to think more clearly about context when their own views are challenged. (5) the “quiz” area located in the lower-right part of the window.

4. Experimental design

To evaluate the efficacy of the social semantic tagging-based mobile learning system for constructing prior knowledge, an experiment was conducted on reading activity at a senior high school in Taiwan. 56 students (24 male students and 32 female students) participated in this study. Each class consisted of 28 students. A quasi-experiment was designed by assigning the students in one class to the
experimental group, and the other class to the control group. All students were taught by the same teacher who had more than ten years’ experience of teaching science courses.

For material selection, reading materials in this study all pertain to science topics, and the materials were ensured to be suitable for senior high school students. In this study, the measuring tools included a pre-test, a post-test, and the questionnaire for measuring the students’ learning achievements, motivations, self-efficacy, and socialization. The self-efficiency measure was developed by Pintrich, Smith, Garcia, & McKeachie (1991). The questionnaire for learning attitude measure was developed by Hwang, Yang, & Wang (2013).

Before the learning activity, an orientation was given to introduce the learning environment and the learning tasks. Moreover, the students took the pre-test and the pre-questionnaire. During the learning activity, the two classes were assigned to the control group and the experimental group. The experimental group learned with the social semantic tagging-based learning guidance embedded mobile learning, while the control group learned with the conventional mobile learning without the social semantic tagging-based learning guidance. After the learning activity, the students took the post-test and filled out the pre-questionnaire including learning motivation, self-efficacy and perceived ease of use and usefulness for comparing the learning achievements and the improvements in learning attitude of the two groups.

5. Experimental Results

5.1 Analysis of learning achievement

To evaluate the effectiveness of SSTL, pretest and posttest evaluations were implemented to demonstrate the achievement of learning outcomes. Here, the pre-test results reveal that the mean score of the experiment group was similar to that of the control group (61.61 and 62.86). The t-test result showed that these two groups did not differ significantly (t = .214, p > .05). In other words, before performing the experiment, the pre-test revealed that control and experimental group demonstrated a similar understanding of the learning topics at an alpha level of 0.05.

After participating in the learning activity, the two groups of students took a post-test. The t-test results of the post-test in Table 1 indicate that the experimental group had a higher mean score than the control group. Furthermore, the results show that the learning achievement of the experimental group was significantly better than that of the control group (t = -3.827, p < .05). This implies that the proposed interactive u-learning system based on a social semantic tagging-based learning guidance benefited the students more than the traditional approach.

Table 1. Paired t-test of the learning improvement for the two groups

<table>
<thead>
<tr>
<th>Variable</th>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>S.D.</th>
<th>Std. Error.</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post-test</td>
<td>Control Group</td>
<td>28</td>
<td>59.29</td>
<td>12.0734</td>
<td>2.2817</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Experimental Group</td>
<td>28</td>
<td>67.85</td>
<td>13.3531</td>
<td>2.5254</td>
<td>t = -3.827*</td>
</tr>
</tbody>
</table>

*p < .05

5.2 Analysis of learning motivations

In order to examine the difference in the learning motivation for students before and after participating in the learning activities, the questionnaire is presented with a 5-point Likert scale where ‘5’ means strong agreement or positive feedback and ‘1’ represents high disagreement or negative feedback. Table 2 shows the t-test result of the learning motivation of the two groups. The results show that the students of the experimental group improved toward learning motivation after the learning activity. In the learning motivation questionnaire, it is found the experimental group has significant difference (t = -4.26, p <.01) between pre- and post-questionnaires. In contrast, the t-test results of the control group showed no significant difference, as shown in Table 2. This result revealed that the learning motivation of the students from the experimental group increases after the learning activity.
Table 2. The paired $t$-test result of pre- and post-questionnaire of learning motivation

<table>
<thead>
<tr>
<th>Group</th>
<th>Question</th>
<th>N</th>
<th>Mean</th>
<th>S.D.</th>
<th>$t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental group</td>
<td>Pre-questionnaire</td>
<td>28</td>
<td>3.67</td>
<td>0.47</td>
<td>-4.26**</td>
</tr>
<tr>
<td></td>
<td>Post-questionnaire</td>
<td>28</td>
<td>4.14</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td>Control group</td>
<td>Pre-questionnaire</td>
<td>28</td>
<td>3.71</td>
<td>0.46</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td>Post-questionnaire</td>
<td>28</td>
<td>3.60</td>
<td>0.49</td>
<td></td>
</tr>
</tbody>
</table>

**p<0.01

5.3 Analysis of learning self-efficacy

A seven-point Likert scheme was applied in the pre-test of the self-efficacy. The three sets of values in the one-way ANOVA (Analysis of Variance) test result are provided as follows: the mean value of the test was 4.21 for the SSTL enhanced Science mobile learning group and 4.14 for the Science mobile learning group. According to the results, no significant difference was shown in the self-efficacy between the two groups in the class ($F=1.42$, $p=0.25>0.05$). Based on the analysis above, this study further compared the two sets of values in the self-efficacy before and after learning groups, as shown in Table 3. The results found that the experimental group had significant difference between the pre- and post-test of the self-efficacy ($t=-3.10$, $p<0.05$). On the contrary, there was no significant difference between the pre- and post-test of the self-efficacy in the control group ($t=-1.15$, $p=0.26>0.05$). This indicated that perceived self-efficacy has been significantly improved after learning with the social semantic tagging-based learning system.

Table 3. The paired $t$-test result of exercise self-efficacy of science education

<table>
<thead>
<tr>
<th>Question</th>
<th>N</th>
<th>Mean</th>
<th>S.D.</th>
<th>$t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSTL enhanced Science</td>
<td>Pre-questionnaire</td>
<td>28</td>
<td>4.03</td>
<td>0.42</td>
</tr>
<tr>
<td>mobile learning group</td>
<td>Post-questionnaire</td>
<td>28</td>
<td>4.36</td>
<td>0.48</td>
</tr>
<tr>
<td>Science mobile learning</td>
<td>Pre-questionnaire</td>
<td>28</td>
<td>4.14</td>
<td>0.52</td>
</tr>
<tr>
<td>group</td>
<td>Post-questionnaire</td>
<td>28</td>
<td>4.32</td>
<td>0.47</td>
</tr>
</tbody>
</table>

*p<0.05

5.4 Analysis of Perceived Ease of Use and Usefulness

To better understand the students’ perceptions of the use of the SSTL learning system, this study collected the students’ feedback in terms of “perceived usefulness” and “perceived ease of use”. Results found that most students gave positive feedback concerning the two dimensions of the SSTL learning system. The average ratings for “perceived ease of use” are 3.75 and 3.42 for the experimental group and the control group, respectively; moreover, their average ratings for “perceived usefulness” are 3.85 and 3.14. In comparisons with ratings given by the control group, it should be noted that the students in the experimental group gave higher ratings to “perceived ease of use” and “perceived usefulness”, implying that the students who learned with the SSTL learning system revealed higher degrees of technology acceptance than those who learned with mobile learning system.

In terms of perceived usefulness, the t-test result ($t=4.67$, $p<0.001$) shows significant between the experimental group and the control group. It depicts that the social semantic tagging-based learning guidance approach (SSTL) is more effective than the conventional mobile learning approach. From the students’ interview feedback, most students in the experimental group agreed with the usefulness of the SSTL learning system approach in improving their learning achievements. Moreover, they could learn better by using interactive learning guidance system and the learning system is helpful to their learning of science education.

6. Conclusions and future work
In this study, the impact of using the social semantic tagging-based learning guidance (SSTL) was explored to enhance the positive impact of science learning. The proposed SSTL approach and interactive u-learning system were developed that can provide a richer understanding of how users can more efficiently employ social semantic tagging to enhance the learning experience. For knowledge construction, the proposed approach provides opportunities for students to demonstrate knowledge connections, because tags as annotations can serve as spontaneous behavior in reading (Chan & Pow, 2020). The experimental results showed that the system’s valuable functions for prior knowledge acquisition and reading comprehension. It was also found that the students of the experimental group had significantly improved in their learning motivation and perceptions.

Despite these encouraging experiment results however, there are still difficulties in creating a quality measurement of semantic tagging for tag-based interactive learning environments. One major problem is that tags have issues with both sparseness and noise. Before performing our experiment, this study used several preprocessing techniques to reduce the influence of sparseness, including Porter stemming and stop word. Moreover, the study also proposes a series of tag implementation guides to ensure that students tag meaningful ideas, but the filtering rule still incomplete. These results also point to suggestions and references for the design of efficient mobile-supported learning activities in the future. Additionally, the small sample size of each group is another problem. Therefore, further research with larger sample size will be needed to investigate this methodological concern and its practical applications.

References


Pandemic 2020 and Education: Responding from Kiribati

Faisal BADAR*, Jon MASON & Marebu FLOOD
Charles Darwin University, Australia
*faisal.badar@students.cdu.edu.au
jon.mason@cdu.edu.au
mkflood1982@gmail.com

Abstract: Growth of unprecedented innovation in Education Technology for several decades faced a major challenge with the pandemic of 2020, characterised by disruption and opportunity. While educational services in high-income economies have benefitted from digital learning solutions during the disruption, most low- and middle-income economies have struggled to provide undisrupted access to learning during the pandemic lockdown. Framed by the global context of the pandemic, this paper identifies a global oversupply of advice from academic and industry experts about best practices during such circumstances. In contrast, the capacity of an education system from a low-income economy such as Kiribati, a small island country in the Pacific Ocean, is profiled. Key questions are identified for conducting qualitative research to gather further data for the case-study so that it might inform a related research agenda focused on developing technology-based solutions for out-of-school children in underprivileged contexts.

Keywords: COVID-19, pandemic, low-income economies, education system, educational technologies (EdTech), Kiribati

1. Introduction

COVID-19 is the biggest pandemic and global health crisis the world has faced in the last 100 years. It has also created sudden and unprecedented challenges in education where formal learning has been disrupted in 194 countries affecting 1.57 billion learners (91.3%) of total enrolled learners globally (UNESCO, 2019). Governments have been forced to close schools, colleges and universities for an indefinite period, and learning shifted to home-schooling sometimes without providing clear directives. The pandemic strikes the world when education systems even in advanced countries were not prepared to adapt digital learning opportunities. The situation in underprivileged areas is even worse, with some education systems almost completely deprived of Education Technology-based (EdTech) learning platforms. It is predicted that the hasty global shift towards online education will further exacerbate inequality in the attainment of education around the world (Petrie et al., 2020).

Leading up to this crisis, the world also experienced exponential growth of Educational Technology (EdTech) networks. For high-income countries, many stakeholders are undergoing a crash course in online learning. These alternate technology-based learning opportunities are not only providing a stop-gap solution during the crisis but also stimulating innovative responses. Digital technology is pivotal to this.

This paper presents a case study of the situation in Kiribati, a small country in the Pacific Ocean with minimal technology resources, as an example response of adapting technology to the challenges of underprivileged education systems during the crisis. As such, it represents a useful reference to a broader issue facing low- and middle-income countries: the provision of learning services to out-of-school children (OOSC). The paper initially outlines the impact of COVID-19 on the education sector worldwide, global efforts to look for solutions for going forward, and how within a very short period, an oversupply of advice from academic and industry experts occupied the learning arena concerning best practices. In contrast, Kiribati’s capacity to respond is profiled followed by a proposed methodology for probing deeper that challenges and opportunities that might inform the response of the government education system are in Kiribati.
2. Global shift in Education during COVID-19

2.1 COVID-19 – Impact on Education

The COVID-19 pandemic represents a global crisis that has triggered an unprecedented shift in educational practices on a global scale. The key question for all stakeholders is how to continue providing access to formal learning during this disruption. In the words of UNESCO Director-General Audrey Azoulay, “We are entering uncharted territory and working with countries to find hi-tech, low-tech and no-tech solutions to assure the continuity of learning” (UNESCO, 2020). Recent literature in education research highlights the need for global resilience and continuity of formal education. Although children have been found minimally susceptible to COVID-19 schools have been closed worldwide (Abdulamir et al., 2020; Faherty et al., 2018; Germann et al., 2019). Studies are also showing that prolonged school closures and staying at home may bring negative impacts on children’s physical and mental health, and the “psychological impact of quarantine is wide-ranging, substantial and can be long-lasting” (Brooks et al., 2020; Brazendale et al., 2017). Nonetheless, technology represents a key part of the solution and earlier studies also show this (Ash et al., 2014).

2.2 Opportunity in Adversity – EdTech in COVID-19 Era

The positive aspect of COVID-19 is that it brings a great opportunity for EdTech to continue to deliver on innovative education solutions. Already, the pandemic has transformed many learning communities into online learning communities, forcing all stakeholders to embrace online learning. Technological innovations in education over the past two decades have already enabled a transformed education landscape (Mason & Pillay, 2015). Intelligent Digital Systems can now efficiently adapt the learning experience to suit personal learning preferences, often with better precision than any traditional classroom can. Similarly, virtual laboratories provide an opportunity to practically design, conduct and learn from experiments, rather than just learning about them (Petrie et al., 2020). Such examples, however, are not equally available to learners worldwide.

Due to format limitations, it is not possible to cover all initiatives in this paper; however, we highlight some key global initiatives that may help students, parents, teachers, and other education stakeholders. UNESCO has developed a live portal COVID-19 Education Disruption and Response to present everyday status updates of schools worldwide in addition to other useful information (UNESCO, 2020). The OECD has provided a framework to guide an education response to the pandemic including a 25-point checklist of education response to COVID-19 with 13 priority responses by countries. HundrED (April 2020) published a report captioned Spotlight: Quality Education for all during COVID-19 crises including a repository of hundreds of resource pages, innovations, learning approaches and educational tools created by teachers, organisations and governments for students, parents and teachers to consult for everyday educational activities, ideas, initiatives and platforms (Petrie et al., 2020). They further conducted a survey through 150 stakeholders in education from 31 countries to understand current responses. The survey takeaways include that (i) 87% respondents were concerned that pandemic will increase educational inequality, (ii) only 6% responded that their education system was highly prepared for the pandemic, and (iii) only 17% of respondents believe that education leaders were learning from other countries’ responses.

One of the primary issues emerging from COVID-19 school closures is the shift in the role of parents and guardians. Government decisions to close schools and to shift to an alternate learning mode has come as an abrupt shock for many families as this crises-mode home-schooling has also impacted their regular productivity in their jobs. Another stakeholder group impacted dramatically are the teachers who are, in most cases, not skilled enough to continue teaching in online mode. A survey conducted by the OECD including 330 responses from 98 countries explored how countries are responding to the pandemic (Reimers et al., 2020). Stakeholder responses included schoolteachers, principals and other staff, civil society organisations, government advisors and policymakers, staff in international organisations, and education consultants. Most respondents highlighted that while governments issued directives for staff and teachers no specific instructions for continuity of education...
during closure and no prioritisation of curriculum and resources was done. Table 1 summarises perceived stakeholders’ priorities while Table 2 shows challenges in implementation.

Table 1: *How critical are the education priorities in response to the crisis*

<table>
<thead>
<tr>
<th>Priority</th>
<th>Did not respond</th>
<th>Not very critical</th>
<th>Somewhat critical</th>
<th>Very critical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ensure the continuity of academic learning for students</td>
<td>145</td>
<td>137</td>
<td>135</td>
<td>131</td>
</tr>
<tr>
<td>Provide professional support, advice to teachers</td>
<td>136</td>
<td>134</td>
<td>132</td>
<td>130</td>
</tr>
<tr>
<td>Ensure well-being of teachers</td>
<td>132</td>
<td>130</td>
<td>128</td>
<td>125</td>
</tr>
<tr>
<td>Support students who lack skills for independent study</td>
<td>131</td>
<td>129</td>
<td>127</td>
<td>125</td>
</tr>
<tr>
<td>Ensure well-being of students</td>
<td>130</td>
<td>128</td>
<td>126</td>
<td>124</td>
</tr>
<tr>
<td>Ensure support for parents and caregivers to support</td>
<td>129</td>
<td>127</td>
<td>125</td>
<td>123</td>
</tr>
<tr>
<td>Ensure continuity/integrity of the assessment of students</td>
<td>128</td>
<td>126</td>
<td>124</td>
<td>122</td>
</tr>
<tr>
<td>Revise graduation/policy transition policy to other</td>
<td>127</td>
<td>125</td>
<td>123</td>
<td>121</td>
</tr>
<tr>
<td>Ensure medical attention to teachers affected by Covid-19</td>
<td>126</td>
<td>124</td>
<td>122</td>
<td>120</td>
</tr>
<tr>
<td>Define new curricular priorities for the crisis</td>
<td>125</td>
<td>123</td>
<td>121</td>
<td>119</td>
</tr>
<tr>
<td>Ensure medical attention to students affected by Covid-19</td>
<td>124</td>
<td>122</td>
<td>120</td>
<td>118</td>
</tr>
<tr>
<td>Ensure provision of other social services to students</td>
<td>123</td>
<td>121</td>
<td>119</td>
<td>117</td>
</tr>
<tr>
<td>Ensure distribution of books to students</td>
<td>122</td>
<td>120</td>
<td>118</td>
<td>116</td>
</tr>
<tr>
<td>Other, specify</td>
<td>121</td>
<td>119</td>
<td>117</td>
<td>115</td>
</tr>
</tbody>
</table>


Table 2: *How challenging has it been to implement the education response*

<table>
<thead>
<tr>
<th>Challenge</th>
<th>No response</th>
<th>No challenge at all</th>
<th>Some challenges</th>
<th>Many challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Availability of technological infrastructure</td>
<td>130</td>
<td>128</td>
<td>126</td>
<td>124</td>
</tr>
<tr>
<td>Addressing students emotional health</td>
<td>129</td>
<td>127</td>
<td>125</td>
<td>123</td>
</tr>
<tr>
<td>Achieving the right balance between digital and synchronous activities</td>
<td>128</td>
<td>126</td>
<td>124</td>
<td>122</td>
</tr>
<tr>
<td>Lack of availability of parents/guardians to support learning at home</td>
<td>127</td>
<td>125</td>
<td>123</td>
<td>121</td>
</tr>
<tr>
<td>Management of technological infrastructure</td>
<td>126</td>
<td>124</td>
<td>122</td>
<td>120</td>
</tr>
<tr>
<td>Lack of adequate communication with parents to co-ordinate curricula-aligned learning</td>
<td>125</td>
<td>123</td>
<td>121</td>
<td>119</td>
</tr>
<tr>
<td>Lack of capacity or willingness of teachers to adopt to the changes imposed by the situation</td>
<td>124</td>
<td>122</td>
<td>120</td>
<td>118</td>
</tr>
<tr>
<td>Other, specify</td>
<td>123</td>
<td>121</td>
<td>119</td>
<td>117</td>
</tr>
</tbody>
</table>


UNESCO (May, 2020) has also published a comprehensive handbook guiding the use of Open Educational Practices (OEP) during school closure and how to utilise Open Education Resources (OER) under COVID-19. OER and OEP are inter-related but different terms, and both leverage the basic open architecture of the internet itself (Mason & Pillay, 2015). Figure 1 illustrates the shift in the focus of researchers and educators from creating and publishing OER to practices that can be implemented using OER, referred to as OEP.
The handbook discusses the use of OEP and OER during COVID-19 through vivid stories and experiences, OER competencies for OEP, and provides guidelines to both teachers and learners to facilitate OEP and OER application. These guidelines are based on five UNESCO (2019) objectives namely: (i) Building the capacity of stakeholders to create access, re-use, adapt and redistribute OER; (ii) Developing supportive policy; (iii) Encouraging inclusive and equitable quality OER; (iv) Nurturing the creation of sustainability models for OER; and (v), Facilitating international cooperation. The handbook provides a comprehensive review of OEP definitions and based on these definitions, the five settings were identified for the applications of OEP in education, i.e. use of OER learning material, open teaching, open collaboration, open assessments, and enabling technologies (Huang, Tlili, et al., 2020). The inter-relationship of these settings is shown in Figure 2.

The handbook further discusses challenges in applying OEP, OER competencies for applying OEP, OER-enabled distance learning strategies and guidelines for applying OEP.

Another joint effort by UNESCO, International Research and Training Centre for Rural Education, and Smart Learning Institute of Beijing University is the Handbook of Flexible Learning during Educational Disruption (SLIBNU, March 2020), focused on the Chinese experience in maintaining undisrupted learning in COVID-19 Outbreak. The handbook re-conceptualizes flexible pedagogy as a learner-centred educational strategy, which provides choices from the main dimensions of study, such as time and location of learning, resources for teaching and learning, instructional approaches, learning activities, support for teachers and learners. It further describes online learning under the theme “Disrupted Classes, Undisrupted Learning” based on seven factors, (i) reliable communication infrastructure, (ii) suitable digital learning resources, (iii) friendly learning tools, (iv) effective learning methods, (v) instructional organizations, (vi) effective support services for teachers and learners, and, (vii) close cooperation between Governments, Enterprises and Schools (G-E-S cooperation). Several flexible online learning strategies are highlighted based on six dimensions, namely (a) infrastructure, (b) learning tools, (c) learning resources, (d) teaching and learning methods, (e) services for teachers and students, and (f) cooperation between enterprise, government, and schools, that can help other educators, researchers and practitioners implement similar case studies in their context. The following discussion on Kiribati places all this guidance in a real-world context of a proposed in-depth case-study.

3. COVID-19 response from Low-income Economies: Kiribati Case-study

3.1 Kiribati: Demographics and Education System

The Republic of Kiribati is one of the smallest island nations in the world but is also considered to be the largest atoll country in the world with a total land area of 811 square km shared across approximately 33 low-lying coral atolls spreading across 3.5 million square km in the Central Pacific Ocean, of which
only 20 are inhabited. Kiribati has a population of 110,136 (in census 2015). About 90% of that population lived in the Gilbert Islands, with 40% of them on South Tarawa.

Kiribati has a basic literacy rate of over 90% (State University, 2020). The basic education system in Kiribati comprises of primary school, junior secondary schools and senior secondary schools. There are 110 primary, 24 junior secondary, 18 senior secondary and 9 combined secondary schools. The government provides supplementary support to lower-level students from Years 1 to 9, and those students in senior secondary levels mostly achieve the government’s benchmarks (Kiribati Government, 2016).

Major problems in the Kiribati education sector include (i) insufficient places in senior secondary schools (ii) high rates of dropouts due to lack of government financial support and limited schools, (iii) lack of resources to provide quality education and qualified educators, (iv) inadequate English skills compared to other Pacific Islanders, and (v) lack of ICT integration in the education system. These problems hinder the progress of education in Kiribati as well as increasing the poverty rate in terms of unemployment.

3.2 Future Research

In probing the situation in Kiribati for further detail we propose a qualitative research study to analyze the education system response due to COVID-19, including the actions taken by the stakeholders, challenges faced during the implementation phase, responses from other stakeholders, and outcomes of their actions. The following research questions provide the focus:

RQ1: What measure has the Kiribati Education System put in place to minimise the impact of COVID-19 on student learning?

RQ2: What has been the readiness of the Kiribati Government and other stakeholders in utilising digital technology for continuity of education during the lockdown and what challenges were faced during implementation?

RQ3: What measures can be taken to continue education in Kiribati to be ready for the reoccurrence of any such situation in future?

Data will be collected via interviews with the stakeholders, such as government officials from the Education Department, School Principals and Teachers, Church authorities involved in educational services and parents of the school-going children. Around 12-15 respondents will be approached through a snowballing technique, and a semi-structured interview will be conducted to gather information from relevant stakeholders.

4. Discussion and Conclusion

This short paper has emphasized the capacity of Kiribati to respond to the crisis of COVID-19. Importantly, we note that while high-income economies have generally been able to respond through deploying digital technology solutions, low-income economies like Kiribati have not yet demonstrated any commensurate recovery. Children in many low- and middle-income countries have been left behind as their governments and education systems could not adequately respond to such a severe lockdown.

With over 90% of learners affected worldwide, COVID-19 has impacted the world immensely and has broadened the gap between have-s and have-nots in education globally. The education deprived majority of the world has been pushed further away from their counterpart who could continue their learning, maybe partially, even during the biggest learning obstruction of the century. Our preliminary research is aimed at informing policymakers, educators and EdTech designers to ensure that underprivileged communities during the COVID-19 pandemic are not left behind as the global economy recovers. The paper may also provide better visibility to the EdTech industry regarding the needs and limitations of these communities and may help them design adaptable, sustainable, and scalable solutions that are suited to the local contexts of these underprivileged learning communities.

In the Kiribati context, where availability and affordability of digital infrastructure (such as computers and connectivity) is a serious issue for individual families, and practical implementation of ICT teaching and learning is a weak area, we suggest formation of multiple Digital Learning Centres (DLCs) at various locations as substitute learning platforms, with the following objectives:
- DLCs be managed under the government or community control, with password-protected time-bound access to all teachers and students during after-school timings with computer usage restricted to assigned work only.
- Regular compulsory ICT-related and practice-focused training for teachers and students, with reflection on learning outcomes in classroom settings be included in career milestones of teachers and academic grades of students.
- Establishment of a digital learning and assessment portal utilising local learning content and OERs to provide a viable alternate learning platform to enhance learning capacity of students.
- Design and development of inquiry and project-based assignments for students, which may be accessed and solved through computers in DLCs.

References


Design Explorations to Support Learner’s Mental Health using Wearable Device and GOAL application

Taisho KONDOa*, Huiyong LIa, Yuan Yuan YANGa
Rwitajit MAJUMDARB* & Hiroaki OGATAB

a Graduate School of Informatics, Kyoto University, Japan
b Academic Center for Computing and Media Studies, Kyoto University, Japan

*kondo.taisho.47x@st.kyoto-u.ac.jp, dr.rwito@gmail.com

Abstract: Increasing the number of patients with symptoms of depression makes maintaining Mental Health concern in many societies calling for global attention. In children, stress and depression have a detrimental effect on their developmental emotions and physical health. From the perspective of a student, daily stress can affect their learning activities. Hence, to tackle this issue, it is necessary to develop their ability to manage mental health. Self Direction Skills (SDS) can support to develop such an ability. Current wearable devices can acquire physical activity and physiological data to compute stress levels, and sleep quality. While it makes it possible to visualize the amount of stress received in daily life, no such program is linked to the daily learning life of a student. To fill in this gap we extend our earlier developed GOAL system which already synthesizes students’ physical activity and learning data and explores possibilities of integrating internet-based Cognitive Behavioral Therapy (iCBT) with our previously proposed DAPER model as an applicable method to deal with daily stress in academic life. We use the wearable devices to automatically track stress levels, and map activities of training students’ mental health managing ability to the DAPER phases. Students can make a plan in the GOAL system to reduce stress by selecting a cycle of practice as suggested by iCBT. To inform our system design we conducted a pilot study with 120 students at a junior high school level. This paper presents an exploratory data analysis of 60 students, whose learning performance from english and mathematics courses and stress levels over 133 days were collated.

Keywords: GOAL, Mental Health, Garmin, Self-direction skills, Automated measurement

1. Introduction

Japan faces an increasing number of individuals with mental health issues such as depression and high stress. As of 2008, the total number of patients with depression exceeded one million. (Ministry of Health, Labor and Welfare, MHLW, in Japan 2010). Managing mental health is a very important issue globally and Japan also focuses on it (Mental Health Welfare Measures Headquarters in Japan, 2004). Psychological counseling is a common way to attend to mental health issues, and in Japan elementary and junior high schools have full-time school counselors as staff to support the students. However, individual attention is difficult to scale for such support. Therefore, we consider it important for students to learn how to analyze their mental health, monitor it, and reflect on a healthy state. Various mental health disorders are often the result of daily stress, closely related to sleep and physical activity (Åkerstedt et al. 2007). Wearable devices such as smartwatch compute stress levels based on the individual’s physiological parameters like heart-rate. It also records physical activity and sleep-related data. Thus, it becomes possible to measure students' mental state to a greater extent with quantitative data automatically acquired by such wearable devices. We had earlier created a platform called GOAL (Goal Oriented Active Learner) system to support the
integration of multiple data sources such as data from learning platforms as well as from activity trackers (Majumdar et al., 2018). The primary objective of GOAL is to support students' acquisition of self-direction skills (SDS) by presenting their log data. In this study, we explore possibilities of extending the GOAL platform to train school children with learning how to manage their stress based on quantitative data gathered from wearable devices. It aims to reduce stress by changing their daily behavior. As a pilot study, we conducted an exploratory data analysis to investigate the stress patterns of students at a junior-high-school level.

2. Related Work

Our method advocates automatic data gathering and skill measurement by the system. In this section, we summarize previous studies (see Table 1) that use log data to support mental health. For example, the pattern of spending money on gambling can be used to determine whether the type is easy to waste money. (Auer, Michael, et al. 2017). CrossCheck (Wang, Rui, et al. 2016) uses passive smartphone sensor data (sleep, mobility, conversations) to build inference models capable of accurately predicting aggregated scores of mental health indicators in schizophrenia. However, none of the support methods has improved the way students approach mental health. Many studies use automatic data gathering from self-report to smartphone but seldom do they provide system support for improving the status of the user. To fill the gap, this paper proposes a technology-enhanced approach to support stress management for students. By visualizing the current status to analyze, then planning, and completing a series of actions in the system, the user is expected to grasp and improve their mental health.

Table 1. Examples of existing technology for mental health support

<table>
<thead>
<tr>
<th>Digital Application</th>
<th>Types of data processed</th>
<th>Usage of data</th>
<th>Purpose</th>
<th>SDS for mental health</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personalized feedback (Auer, et al. 2017)</td>
<td>The amount of money lost in the gamble</td>
<td>model using machine learning</td>
<td>Solve cognitive dissonance</td>
<td>×</td>
</tr>
<tr>
<td>CROSS CHECK (Wang et al. 2016)</td>
<td>passive smartphone sensor data (sleep, mobility, conversations)</td>
<td>track current mental status, and model using machine learning</td>
<td>Early detection of mental health changes for patients with schizophrenia</td>
<td>×</td>
</tr>
<tr>
<td>GOAL for mental health (based on Majumdar et al. 2018)</td>
<td>The stress level from the wearable device</td>
<td>Quantification and visualization of current mental state</td>
<td>Improve Self-directed mental health monitoring</td>
<td>○</td>
</tr>
</tbody>
</table>

3. GOAL for self-directed mental health support

3.1 Tasks for mental health activity based on DAPER model

Adapting DAPER model, we show the process of improving SDL skills for mental health using the stress data obtained from the wearable device. The DAPER model consists of five phases. Data Collection, Data Analysis, Planning, Execution Monitoring, Reflecting. In the data collection phase, students synchronize activity, sleep, stress level from their wearable devices with the GOAL system. For these data, the values sent from the devices every day are
aggregated every day and displayed as a daily average value of one student. In the analysis phase, they can conduct simple analysis tasks to understand their activity trends and identify if there are any issues. Making a SMART (Specific, Measurable, Appropriate, Relevant, and Timely) plan would be the next step. Students can choose to make a plan of step count, sleeping time, or stress level to develop good health habits, and the system uses the average value of a group for comparison. There are three types of tasks: activity, sleep, and iCBT practice. Activity and sleep are compared with the average value of 1 month per class based on the data that can be measured by a wearable device and objectively judged whether the value is small. We also suggest iCBT intervention. In the execution phase, the individual monitors the progress of their plan and in the reflection phase, they review the whole process. It involves evaluating the difficulty of each task and the effort they made to achieve the goal in their chosen activities.

4. Pilot Study for data exploration

In the pilot study, we investigated the distribution of the stress level as collected from the wearable device and its relationship between learning performance behavior. With the understanding of the stress level variation and its relationship to different performance cohorts, can potentially help to decide the support strategy in our future work.

4.1 Context and participants

We conducted a pilot data collection in a public school in Japan. Garmin Vivosmart3 smartwatch was distributed to 120 students from grade 3 junior high school. Since some students did not wear it or synchronize their data, there was stress data collected in the GOAL system for 60 students. Performance scores of 94 students were collected, and final target students were 60 (21 Males, 39 Females) students. The project was approved by the city school counsel for the pilot study and the GOAL platform pseudonymously connected students’ Garmin data. Based on that collected data we analyze the following two research questions:

RQ1: What are the weekly stress patterns for Junior high school students?
RQ2: What are the differences in the stress patterns for different groups of students based on their performance transitions?

4.2 Dataset and Analysis method

The data collection period was from November 21, 2019, to April 2, 2020. Over these 99 days, 4054 physical activity data, and 1426 stress data were synchronized in the GOAL server from 60 learners.

The stress level ranges on a scale from 0 to 100 and automatically computed by the Garmin device based on the heart rate measurement (Heart Rate Variability and Stress Level, 2020). It classifies resting state (0 to 25), low stress (26 to 50), medium stress (51 to 75), and high stress (76 to 100) state. Stress duration is the number of seconds in this monitoring period where stress level measurements were in the stressful range (26-100). Daily average, maximum, and total stress values of each user are synchronized in the GOAL server. To answer RQ1 we compute the day-wise distribution of the average stress values and the duration for each day of the week across the period.

Besides the stress data, we also collected students’ learning performance data in Mathematics and English subjects across the semester. It consists of two term-tests conducted during 26-27 November 2019 and 25-27 February 2020. To answer RQ2 we use the Stratified
Attribute Tracking (iSAT) approach (Majumdar, Alse & Iyer 2014) to first divide students into four groups based on the changes of their grades over the two-term period. For each of the English and Mathematics subjects respectively, four groups would indicate performance remaining at high or low or improving or deteriorating. Sixteen transition groups were generated considering both English and Mathematics performance. Figure 2a shows the transition patterns of group memberships across the two subjects. From this distribution, we were interested in investigating the difference between students who belonged to the three highlight transition groups: both low performers (n=12), both high performers (n=12), mixed (n=36). Applying the Shapiro-Wilk test, we found the Average Stress level data were normally distributed (p=0.119), but Stress duration was not normally distributed (p=0.0005). Hence ANOVA and Kruskal-Wallis tests respectively were chosen to determine whether Stress levels and Stress duration were significantly different in the three groups.

4.3 Results

4.3.1 Weekly stress patterns

To answer RQ1, Stress level, Stress duration by day of the week is presented as a box plot in Figures 1a and 1b respectively. In average stress levels, the average second quartile is in the range of 10 to 30 on every day of the week, so half of the students are less stressed. Comparing the stress values of the students who feel the most stress on any day of the week with the stress values of the second quartile, there is a difference of at least 20 and it can be seen that the stress values. In stress duration, it can be seen that the difference in maximum stress time is about 140 minutes, which varies greatly depending on the student.

![Figure 1. Distribution of average a. stress level across days of week. b. stress duration (minutes)](image)

4.3.2 Differences in the stress levels of performance cohorts

Results indicate no significant difference between stress levels of high performance in both subject group (M = 22.6, SD = 14.3), low performance in both subject group (M = 21.0, SD = 12.3) and mixed group (M = 18.6, SD = 11.7), p=0.687. Further, there were no significant difference between
stress durations (minutes) of High group (M = 208, SD = 211), Low group (M = 122, SD = 173) and mixed group (M = 130, SD = 160), p = 0.772.

4.4 Findings from the dataset

In the results of this pilot study, there was no relationship between stress and performance. However, from Figure 2b and 2c, it can be seen that the highest performing group receives the highest third quartile of stress. Therefore, it can be inferred that a certain amount of stress is associated with improving the grade.

5. Discussion and Conclusion

In this paper, we proposed an approach to use wearable technology and GOAL platform to track and support students' self-directed mental health monitoring. In the current times of the COVID-19 pandemic, the situation is even more restrictive and the number of children who were unable to have social interactions is increasing. Self-learning and developing the ability to control one's health can be a very important ability in these uncertain times. An activity based on the GOAL system might prove effective, as learners can use the system and a smartwatch to start their learning cycle. While the pilot data collection focused on familiarising the students with the smartwatches, no specific interventions were given related to reflecting on mental health status. It was also not compulsory for the students to continue wearing watches. Hence only half of the students used the watches and connected to the GOAL system. While the data is synchronized, the reliability is constrained by the physical sensing capacity of the smartwatch and its algorithm used, whose validity we just assumed. The GOAL system enables continuous learning, so we can expect further stress reduction by introducing a more effective iCBT mechanism. Still, due to irregular usage of the smartwatches and the GOAL system, the current dataset might not have shown any strong relations among its various attributes. A more controlled study is planned for the next semester. The device is distributed to all junior high school students as well as some high school students.
Acknowledgments
This research was supported by the NEDO Special Innovation Program on AI and Big Data 18102059-0, JSPS KAKENHI 16H06304, 20K20131, SPIRITS 2020 of Kyoto University.

References
https://www.e-stat.go.jp/stat-search/database?page=1&query=%E6%82%AA%E6%80%A7%E6%96%B0%E7 %94%9F%E7%89%A9%20%E4%B9%B3%E6%88%BF&layout=dataset&statdisp_id=0003013573
Remi Noguchi, 2016 Does Online Five Minute Cognitive Behavioral Therapy and Emotionally Acceptable Mindfulness Exercise Reduce Depressive Symptoms? - Veriﬁcation by randomized controlled trial(オンラインによる5分間認知行動療法と感情を受け入れるだけのマインドフルネス・エクササイズはうつ症状を軽減するか？－ランダム化比較試験による検証)
Torbjörn Akerstedt, Göran Kecklund, John Axelsson,BiologicalPsychology, Impaired sleep after bedtime stress and worries, Volume 76, Issue 3, 2007, Pages 170-173, ISSN 0301-0511
Wang, Rui et al. (2016). CrossCheck: toward passive sensing and detection of mental health changes in people with schizophrenia. 886-897. 10.1145/2971648.2971740.
Construction of Reasonable Accommodation by Chatbot Using SNS and Operation of an Accessibility Center

Noboru UENO*ª¹, Hiroyuki MITSUHARA® & Masami SHISHIBORI®
ª Graduation School of Advanced Technology and Science, Tokushima University, Japan
ª Faculty of Human Life Science, Shikoku University, Japan
® Graduation School of Technology, Industrial and Social Sciences, Tokushima University, Japan
* c501947004@tokushima-u.ac.jp

Abstract: At Shikoku University, Japan, reasonable accommodation was introduced approximately four years ago for students who applied for consideration. We conducted a questionnaire survey querying students to ascertain their preferred communication method. As a result, we find that most students want an easy way to contact faculty using the LINE SNS app. We are proposing a system that uses AI chatbots via LINE for students who have been given reasonable accommodation.

Keywords: reasonable accommodation, learning support, AI chatbot, SNS

1. Introduction

Figure 1 shows the flow of correspondence between students who need academic support, especially reasonable accommodation, and related faculty members, from the time they enter university to graduation. This correspondence can be classified roughly into three categories, as follows:
(a) Anonymised questionnaire and study support: after enrollment, depending on the result of the student questionnaire regarding their concerns, other factors.
(b) Application procedure for the provision of reasonable accommodation: the application for consideration, the interview, review/approval, planning, agreement/signature, etc.
(c) Daily consultation and provision of information after consideration has been provided: in addition to consulting with students and dealing with their queries, etc.

It is extremely difficult for faculty and staff to provide all this support due to time constraints. A questionnaire survey entitled “About the portal site for consideration only” was conducted with five students from the department. They had already been provided with reasonable accommodation for current accessibility on the university portal site and homepage.
2. Results of the Questionnaire on Portal Site Plan Dedicated to Consideration

In this questionnaire, we asked 11 questions anonymously from the web-only questionnaire page. All respondents were female, and there were five students with disabilities. Namely, three students had mental health disabilities, one a mental and hearing disability, and one was deaf. In line with other research results, the proportion of women with mental health disabilities is much higher than that of men [1]. Table 1 shows each question and the answers provided.

<table>
<thead>
<tr>
<th>Questions</th>
<th>Student A</th>
<th>Student B</th>
<th>Student C</th>
<th>Student D</th>
<th>Student E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1: What is your preferred way to contact university faculty and staff members? (Multiple selections possible)</td>
<td>Portal site e-mail</td>
<td>e-mail</td>
<td>SNS such as LINE e-mail</td>
<td>SNS such as LINE e-mail</td>
<td>SNS such as LINE e-mail</td>
</tr>
<tr>
<td>Q2: Do you need to ask questions about university life, such as about lectures, on the site?</td>
<td>Necessary</td>
<td>Necessary</td>
<td>Necessary</td>
<td>Necessary</td>
<td>Necessary</td>
</tr>
<tr>
<td>Q3: On that site, do you want to create new applications, make changes, or suspend reasonable consideration?</td>
<td>Necessary</td>
<td>Unnecessary</td>
<td>Necessary</td>
<td>Necessary</td>
<td>Necessary</td>
</tr>
<tr>
<td>Q4: Do you wish to check the contents of the current consideration plan on the site?</td>
<td>Necessary</td>
<td>Unnecessary</td>
<td>Necessary</td>
<td>Necessary</td>
<td>Necessary</td>
</tr>
<tr>
<td>Q5: Do you wish to check today's lecture schedule via a listing on the site?</td>
<td>Don't know</td>
<td>Necessary</td>
<td>Necessary</td>
<td>Necessary</td>
<td>Necessary</td>
</tr>
<tr>
<td>Q6: Do you wish to pre-book classroom seat selection on the site?</td>
<td>Necessary</td>
<td>Unnecessary</td>
<td>Don't know</td>
<td>Don't know</td>
<td>Necessary</td>
</tr>
<tr>
<td>Q7: When you cannot attend a class, would you like to access it online?</td>
<td>Necessary</td>
<td>Necessary</td>
<td>Don't know</td>
<td>Necessary</td>
<td>Don’t know</td>
</tr>
<tr>
<td>Q8: When you take a break from a lecture, do you want to check alternative assignments on the site and upload submissions?</td>
<td>Necessary</td>
<td>Necessary</td>
<td>Necessary</td>
<td>Necessary</td>
<td>Necessary</td>
</tr>
<tr>
<td>Q9: Have you any other requests for consideration that should be added?</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Anonymous consultation</td>
</tr>
<tr>
<td>Q10: If you use such a site, do you have any concerns?</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Q11: What do you think is the most important support to allow you to graduate?</td>
<td>None</td>
<td>Environment where consultation is easy</td>
<td>None</td>
<td>Like other students, I just want them to support me</td>
<td>Support for until graduation &amp; support for job hunting</td>
</tr>
</tbody>
</table>

Concerning the students’ preferred method of contacting the university in Q1, all of the respondents answered that using existing familiar communication methods such as e-mail and SNS was convenient for them and thus, convenience was the most important factor.

Q2 to Q8 relate to the type of system required for the web application planned for the future. Of these, for “general questions relating to issues such as lectures and missed lectures,” the students mainly responded that these functions were needed, and that a counselling service was essential when planning a dedicated portal site.

Q9 also included a question about anonymous consultation as a desired function, so it is advisable to set up a separate anonymous consultation desk for the consulting aspect referred to in Q2.

Regarding concerns in Q10, all of the respondents answered that there is no particular problem.

For Q11, many students attached importance to “consultation and support” as the most relevant prerequisites for graduation. Thus, communication with students is the objective in the dedicated portal site plan and must be reflected in the design.
3. Plan for Conversational Chatbot Using SNS

According to the results of the questionnaire survey, many expressed a desire to use SNS as a way to contact faculty and staff. A number of researchers note the importance of communication using such web applications [2], and we would like to actively promote their use. Conversational chatbots can be used to great effect to compensate for the time constraints involved in making contact with a large number of students [3].

The detailed functions and features of each module will now be explained with reference to Figure 2.

(a) The first step is to offer a web accessibility site that is open to all university students, and its main feature is that anyone can consult anonymously.

(b) The next step is for students to log in to their LINE and LINE @ accounts to enable more detailed procedures, and a series of application procedures.

(c) LINE @ is a student only SNS for those who have already received reasonable consideration, as well as daily questions and counselling based on the consideration plan.

![Figure 2. UI design of the chatbot system using SNS.](image)

4. Conclusion

One of the solutions outlined here is the construction of an appropriate and multi-functional database using AI chatbots in order to determine whether or not reasonable consideration is necessary, depending on the chatbot’s answers and context. To this end, we plan to obtain further necessary data from faculty and staff at Shikoku University via a questionnaire, in order to create a database about study support and reasonable accommodation, and to conduct practical experiments to improve the current system so that it can be implemented in a more efficient way.

Acknowledgement

We wish to express our sincerest thanks to Shikoku University for their cooperation with this study.

References


Predicting Stag and Hare Hunting Behaviors Using Hidden Markov Model

Rex BRINGULA* & Ma. Mercedes RODRIGOB
a,b Ateneo de Manila University, Philippines
*rexbringula@gmail.com

Abstract: In this paper, we used Hidden Markov Model (HMM) to describe the gaming behaviors of students and whether they will exhibit “stag” or “hare” hunting behavior in a mobile game for mathematics learning. We found that there is a 99% probability that the students will stay either as stag or hare hunters. Our results also suggest that they would choose arithmetic problems involving addition. These game behaviors are not beneficial to learning because they are only exhibiting mathematical skills they already know. The results of the study show that stag and hare hunters have unique traits that separate the one from the other.

Keywords: collaborative learning, game behavior, mathematics, mobile learning, mobile games

1. Introduction

A collaborative learning environment posits that students learn while they are in a social group (Gillies, 2016). As social learners, students may display varying degrees of participation (Bringula et al., 2018). Stag and hare hunting behaviors are the tendency of a learner to contribute to the scores of the group by choosing either a high-risk mode but with higher points or a low-risk mode but with lesser points (Bringula & Rodrigo, 2019). This game behavior is relatively unknown in the field of a collaborative mobile learning environment. In this study, we attempted to describe the game behaviors of the students and predict whether students will exhibit the stag and hunting behaviors. The output of the study will serve as a basis in the development of adapted mobile-based learning for mathematics.

2. Method

This study used the dataset collected by Bringula and Rodrigo (2019). The dataset is composed of interaction log files that were generated through the use of a mobile collaborative game named Ibigkas!Math. It is a mobile-based learning application for grades 1 to 6 students. It is a collaborative game that covers arithmetic problems (addition, subtraction, multiplication, and division of whole numbers and fractions). The students utilized the mobile game for 15 minutes in an experiment conducted during their class session. The application generates arithmetic problems and it displays in one of the team members’ mobile devices. The player has to read aloud the arithmetic problems. The answers are presented in multiple-choice. The correct answer will appear in one of the team members’ devices. The complete description of this game is discussed extensively in Bringula and Rodrigo (2019).

The dataset contains 4,628 solved problems. The log files have six features, such as difficulty level, type of problem solved, speed, time spent, and the number of correct attempts. The difficulty level and type of problem solved are game modes. In the difficulty level, students may choose a very easy (ve), easy (ea), medium (md), hard (ha), or very hard (vh) game setting. The type of problem solved is the type of arithmetic problems (i.e., addition (add), subtraction (sub), multiplication (mul), division (div)) that the student attempted to solve. Students were informed that prizes would be awarded to three groups of students with the highest game scores. Twenty-five boys and 12 girls with an average of 11 years old participated in the study. They had varying degrees of mathematics abilities – eight were low-performing, 13 were average-performing, and 16 were high-performing.
Speed is a game setting that determines the pace of the game. Students may choose a very slow (2 points), slow (5 points), medium (10 points), fast (15 points), or very fast (20 points) mode. Students were informed that there were no deductions of points for wrong answers but on the time limit. Students that chose the first-three settings were labeled as hare hunters; otherwise, they were labeled as stag hunters. We used the hidden Markov model (HMM) to predict whether students would exhibit a stag or hunting behavior given the sequence of difficulty level and type of problem solved. These variables were used because we found in our separate study that these were the significant variables of a decision tree model that could classify stag and hare hunter behaviors (Bringula & Rodrigo, 2020). There are 20 possible game settings (or states) (5 difficulty problems times 4 types of problem) that a participant may choose from. Furthermore, there are 400 possible transitional states (20 states multiplied by 20 states). Initial probabilities for the HMM were determined based on the log files. HMM using the Viterbi algorithm was implemented in Jupyter Notebook. Ten random samples with ten random observations were fed into the HMM model to determine the possible game behavior of the students. The number of random observations is based on the average number of questions a student can solve per game session. On average, one student solved nine questions per game session.

3. Results and Discussion

We found that there is a very high probability (99%) that the students will stay at their current game behavior (Figure 1). The result suggests that the students will persistently choose a game mode based on the speed setting. This means that they will choose a game setting that they are comfortable with and are less likely to explore new game speed settings. This game behavior may not be beneficial to mathematics learning since students are only exhibiting the skills that they already know.

The graph in Figure 1 also shows that they choose only 10 states out of the 20 possible states. These states are easy and addition (eaadd), easy and multiplication (eamul), easy and division (eadiv), very easy and subtraction (vesub), very easy and addition (veadd), very hard and addition (vhadd), medium and addition (mdadd), medium and multiplication (mdmul), medium and subtraction (mdsub), and hard and addition (haadd). It is worth noting that they are more likely to choose problems involving addition. These game behaviors can be attributed to the fact that the goal of the students is to achieve high scores. It can be also observed from the figure that the distribution probabilities in hare hunters are scattered among the 10 states. The finding implies that hare hunters are solving diverse problems than stag hunters.

Figure 1. A hidden Markov model for relating the game mode settings to the game behavior (stag and hare)

Table 1 shows the three possible cases the students may exhibit during game sessions. The other seven possible cases have the same output as the first observation. It indicates that students are more inclined to be hare hunters – a "slowly but surely" game behavior. The results shown in Table 1 further confirm that hare hunters solve a variety of problems. The hare hunters will attempt all arithmetic problems with varying levels of difficulties.

Meanwhile, a stag hunter tends to solve very hard addition problems. Perhaps, they believe that they will gain more points by solving more difficult problems. Consistent with the definition of stag behavior, the students who chose to set the game mode in a faster setting are risk-takers. A faster game setting may incur higher points but the game imposes higher time deductions for every wrong answer. The HMM model can also predict that students may also shift from being a stag to a hare hunter (and vice versa). As previously shown, the shift in game behavior is very minimal. Nonetheless, it provides an opportunity for learners to explore more game settings.
### 4. Conclusions and Future Work

This study aims to describe the game behavior of students who utilized a mobile game learning application in mathematics and predict whether they will exhibit a stag or a hare hunter game behavior. It is found that there is a very high probability that students will stay at the game settings they choose until the end of the game. Students that display a hare hunting behavior solve a variety of problems in an easy and slow pace manner. This is different from the students who are stag hunters. Stag hunters are risk-takers—solving more difficult problems also with a faster game setting. Thus, we found evidence that stag and hare hunters of a mobile game learning application have unique traits that separate the one from the other.

The results of this study will be implemented in the development of an adaptive version of *Ibigkas! Math*. The adaptive version will be then tested to determine the changes in the gaming behaviors of the students. Its influence on students’ mathematics learning will also be investigated.

### Acknowledgements

The researchers are indebted to Dr. Nieva Discipulo, Ellen Tabayan, UE Elementary and Senior High School Laboratory School staff, my research assistants, and to all participants of the study. This paper is funded by Engineering Research and Development for Technology.

### References


Story Generation System Using Player’s Emotions for Review in Gamed-Based Learning

Shusuke Sato*, Kaoru Sumi**

*School of Systems information of Science, Future University Hakodate, Japan
*g2119018@fun.ac.jp
**kaoru.sumi@acm.org

Abstract: We propose a game-based story generation system that automatically generates scripts in real time by using a player’s emotions and actions. In this paper, we consider whether the history of the experience, containing action and emotion information, would be useful for review of the learning experience. The system provides a player with a virtual world and a virtual tool operated by using a hand controller. The system recognizes the player’s real-time emotions through facial expressions, and it outputs reactions based on these emotions via knowledge-based systems when the player operates the tool. Then, it outputs scripts based on the emotions and the history of actions. We evaluated the system by conducting experiments with university students as subjects. As a result, the system could generate interesting stories based on the player’s experience in the game by using the player’s history of actions and emotions. The results suggest that the historical information, including the learner's real-time actions and emotions, is helpful for review in learning.

Keywords: Game-based learning, story generation, emotions, facial expressions, experience history

1. Introduction

Story generation has typically been performed using text until now (Dehn, 1981) (Eger, Potts, Barot, & Young, 2015) (Pérez, 2015) (Sousa, 2000) (Kaptein & Broekens, 2015) (León, & Gervás, 2014). For example, Dramatica (Anne & Huntley, 2004), a script creation tool, supports creating stories by inputting character dynamics, plot dynamics, and static plot points in a text format. Similarly, Dramatis (Riedl, 2016) infers a reader’s emotions by analyzing the behavior of the characters in a story. Facade is an interactive drama in virtual space (Mateas & Stern, 2003), which can express a player’s attitudes and emotions by text during a story through interaction between the player and the system.

We believe that a livelier story can be created using emotions. Recent years have seen the development of story generation systems using emotions. For example, one system generates a story by changing the emotions preset in non-player characters (NPCs) according to their interactions (Chang, Soo, 2009). It seems, however, that previous studies (Hernandez, Bulitko, & Hilaire, 2014) could not effectively capture real-time emotions, including subconscious emotions.

One previous consumer video game (Booth, 2009) involved the player using a biosensor to obtain vital data on the degree of tension, from which the enemy’s appearance pattern was adjusted accordingly. In addition, other studies (Lopes, Liapis & Yannakakis, 2015) (Togelius & Yannakakis, 2016) (Yannakakis & Togelius, 2011) (Gilroy, Porteous, Charles, & Cavazza, 2012) have captured emotions by using biosensors and user reactions, enabling object placement, mapping, and game design in stage creation. There is almost no previous research, however, on emotion-based automatic story generation designed to capture real-time player emotions during gameplay, with adaptation of the script accordingly.

Because many traditional narrative generation systems focus on content from a third-person viewpoint and do not consider emotion-based stories from a first-person viewpoint, we instead propose a different approach. Specifically, we propose a system that can develop content and generate more personal stories by using the user’s own real-time emotions and actions from a first-person viewpoint. The system is novel in that it can capture and use real-time emotions.
The proposed system outputs the user's game experience as a text history. In this paper, we consider this game as a learning experience accompanied by real-time emotions such as joy and surprise. The system used in this paper was developed as a prototype system for experiments to investigate the effect of user's emotional and action history on game-based learning. Game-based learning (Plass, Mayer & Homer, 2020) (Kapp 2012) is a method of learning while playing a game. Children nurture sociality and cognitive and emotional development through various games such as hide-and-seek and playing tag. In this game, we provide users with the learning experience to use the magic tool that can extract the contents of any object in the virtual space. We predict that a learner's later review of this history, after a learning experience through the game, will be a learning aid. Therefore, in this paper, we examine whether the history of the learner's actions and emotions in real time during learning helps the learner's review.

2. Game-Based Story Generation System Using Player's Emotions

The proposed system, shown in Figure 1, provides game-based story generation using the player’s real-time emotions and actions. It performs emotional recognition in real time by using a webcam to acquire facial images of the player. Specifically, the player uses the Oculus Touch, and the system classifies the player’s actions by tracking the movement of his or her hand. We focus on emotion and behavior because previous studies have shown that body movement conveys emotion (Berthouze & Isbister, 2016).

The virtual space in the proposed system resembles a closed space found in everyday life, consisting of an office and a break room. The office and break room are not separated by a wall, so the player can move freely between the two areas. Figure 2 shows a bird's-eye view of the virtual space. The office has a desk and a copy machine, and the break room has a vending machine.

In the virtual space, the player can act upon objects of interest. To have players experience phenomena that do not occur in reality, they can use a “magical” tool that can take transparent objects from among a set of target objects. The magical tool consists of a black circle of paper, which is used to take objects by touching the desk or vending machine. Table 1 lists these interactive objects. For example, if the player touches the desk, he or she will take a document, a book, or a snack according to the player’s emotion, such as “fear,” “anger,” or “joy.”

![Figure 1. Overview of the proposed story generation system.](image1)

![Figure 2. Bird’s-eye view of the virtual space in the system, which represents an office and a break room.](image2)

<table>
<thead>
<tr>
<th>Object</th>
<th>Internal object</th>
<th>Detected emotion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desk</td>
<td>Document</td>
<td>Fear, Disgust</td>
</tr>
<tr>
<td></td>
<td>Book</td>
<td>Anger, Sadness, Surprise</td>
</tr>
<tr>
<td></td>
<td>Snack</td>
<td>Joy, Contempt, None</td>
</tr>
<tr>
<td>Vending machine</td>
<td>Cola</td>
<td>Fear, Disgust, Sadness, Surprise</td>
</tr>
<tr>
<td></td>
<td>Money</td>
<td>Joy, Anger, Contempt</td>
</tr>
<tr>
<td></td>
<td>Cockroach</td>
<td>None</td>
</tr>
</tbody>
</table>

Table 1. List of interactive objects in the system
The system was inspired by a short movie called “The Black Hole.”¹ In the movie, the black hole was a mysterious sheet of paper printed by a company employee in an office. It enabled human hands to pass through objects and take other objects inside. With cunning, the employee used the strange paper to take out snacks from a vending machine for free. In our system, the player uses the Oculus Touch (Figure 3) to “take out,” “throw,” “put back,” “get,” and “eat/drink”. As shown in Figures 4 and 5, the player operates the tool and experiences emotions at the same time, and then the system generates a story. For example, the player can use the tool on the desk to take out a snack and eat it. Figure 6 shows a situation in which the player takes a cockroach out of the vending machine, is surprised, and throws it away. Figure 7 shows an example of a generated script.

¹ The Black Hole | Future Shorts
https://www.youtube.com/watch?v=P5_Msrdg3Hk
Table 2. Example of a generated story

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>You had been sleeping for some time because you were tired from work.</td>
</tr>
<tr>
<td>2</td>
<td>You woke up at your own workplace with no one around.</td>
</tr>
<tr>
<td>3</td>
<td>You noticed that the copier was moving, and you went to the front of the copier.</td>
</tr>
<tr>
<td>4</td>
<td>A black circle was printed on a piece of paper that emerged from the copier.</td>
</tr>
<tr>
<td>5</td>
<td>You felt a sense of strangeness and touched the black circle.</td>
</tr>
<tr>
<td>6</td>
<td>Suddenly, your hand was sucked into the circle.</td>
</tr>
<tr>
<td>7</td>
<td>You were surprised and removed your hand.</td>
</tr>
<tr>
<td>8</td>
<td>You took a book out of the desk with surprise.</td>
</tr>
<tr>
<td>9</td>
<td>You put the book back with surprise.</td>
</tr>
<tr>
<td>10</td>
<td>You took documents out of the desk with a lack of interest.</td>
</tr>
<tr>
<td>11</td>
<td>You put the documents back with a lack of interest.</td>
</tr>
<tr>
<td>12</td>
<td>You took a book out of the desk with surprise.</td>
</tr>
<tr>
<td>13</td>
<td>You put the book back with surprise.</td>
</tr>
<tr>
<td>14</td>
<td>You were holding the paper in your hand, but you did not feel like doing anything.</td>
</tr>
<tr>
<td>15</td>
<td>You decided to spend the day at the office.</td>
</tr>
<tr>
<td>16</td>
<td>In the morning, the paper was back to its original state, and you told your colleagues what you had experienced.</td>
</tr>
</tbody>
</table>

Table 3. Words generated from player emotions and actions

<table>
<thead>
<tr>
<th>Recognized emotion</th>
<th>Generated word</th>
<th>Action by Oculus Touch</th>
<th>Generated word</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joy</td>
<td>happily</td>
<td>hold trigger and pull hand</td>
<td>took out</td>
</tr>
<tr>
<td>Fear</td>
<td>afraid, nervously</td>
<td>release trigger while shaking controller</td>
<td></td>
</tr>
<tr>
<td>Disgust</td>
<td>disgustingly</td>
<td>release trigger</td>
<td>threw</td>
</tr>
<tr>
<td>Sadness</td>
<td>sadly</td>
<td>bring object to character’s body</td>
<td>put back</td>
</tr>
<tr>
<td>Surprise</td>
<td>with surprise</td>
<td></td>
<td>got</td>
</tr>
<tr>
<td>Contempt</td>
<td>with contempt</td>
<td></td>
<td>ate/drank</td>
</tr>
</tbody>
</table>

Through this content in the system, the player can try an experience that is impossible in real life, by using the magical tool to extract anything inside an object in the virtual space. Thus, players can learn by trial and error in a virtual space through an experience that we do not normally have.

2.1 Motion Recognition

The Oculus Touch is a controller used for motion recognition. It uses two sensors to track its position in a three-dimensional virtual space. As the two sensors respond to the controller’s movements, the virtual space displays a virtual hand that moves in the same way as the player’s actual hand. The Oculus Touch has three input formats: button, analog stick, and trigger. In this system, we did not use a head-mounted display, because the system had to acquire the player’s facial expressions. The system recognizes the action of holding the Oculus Touch trigger. If an object is located at the hand position, the player can hold the object. If the player releases the trigger while holding the object, the virtual hand releases the object.

2.2 Emotion Recognition

The system performs emotion recognition in real time by using Affdex with the Facial Action Coding System (FACS) (Ekman & Friesen, 1977). FACS encodes expressions of emotion by combining action units (AUs), which are the smallest units of visually distinguishable facial expressions. Affdex makes 20 predictions per second. The system uses Affdex and FACS to acquire seven types of emotions: joy, fear, disgust, sadness, anger, surprise, and contempt. We assume that if a special sensor device is used to acquire emotions, the player may be tense and unable to exhibit accurate emotions. Therefore, this system acquires player emotions via a camera.

One previous study (Magdin & Priker, 2017) tested the accuracy of emotion recognition with Affdex. That study found that emotions can be recognized with high accuracy from in front of the face at a distance between 20 cm and 750 cm.
The advantage of using facial information is that the player can use the system freely, without wearing sensors (Kotsia, Zafeiriou, Goudelis, Patras & Karpuzis, 2016). The system outputs scripts according to the player’s emotion and action. It uses 30 predictions of 1.5 s each to determine the correct action. Affdex continuously stores the scores for each of the seven emotion types at a rate of approximately 30 times per second. When the player performs an action, the system sums the scores for each emotion type from the 30 most recent observations. It then determines which emotion was most strongly expressed and was the player’s true emotion.

2.3 Text Generation

The system produces scripts by combining the following: (1) the action performed with the Oculus Touch, (2) the player’s recognized emotion, (3) the object acquired in the virtual space, and (4) the player’s emotion information according to the object. The system has two types of text: one type generated from these four pieces of information, and another type automatically generated by the system. The first type is created by incorporating the four pieces of information into a sentence structure. Rows 8–13 in Table 2 give examples of the generated sentences.

As an example of sentence generation, suppose that the system generated and displayed the sentence shown in Figure 7: “You nervously took the cola out of the vending machine.” The part corresponding to “You nervously” was generated because the system detected the emotion “fear.” Table 3 summarizes the outputs when each emotion is detected. As for the rest of the sentence, the term “vending machine” was the target object of the motion in the virtual space, and “juice” was obtained from knowledge of the vending machine object. Finally, the action “took out” was determined from the acceleration of the Oculus Touch and the part of the character that touched the object in the virtual space. This was used to determine the action result and its output, which was modified according to the emotion. Table 3 also lists the possible actions determined by the system, i.e., taking, throwing, putting, getting, eating, and drinking.

In this system, the story generation follows a dramatic narrative structure. The system provides the introduction and ending by using the given sentences listed in rows 1-7 and 14-16, respectively, in Table 2. The selected story ending differs depending on the player’s actions. Three endings are available: (1) the player obtains three objects, (2) the player eats three foods, or (3) the player does nothing for 5 min.

Thus, the system visualizes and displays the results obtained by the knowledge-based method from the acquired actions and emotions. Here, we explain the sentence generation process using the sentences in Table 2. Rows 1-7 list introduction sentences displayed in the text space at the bottom of the screen so that the player can understand the instructions. For rows 8, 10, and 12, the player used the magical tool to “take out” an object from the desk; therefore, the taken internal object corresponds to the emotion determined from the knowledge of the target external object (i.e., the desk). For rows 9, 11, and 13, the player “put back” the object. The story ends when a specific action is not performed a certain number of times within 5 min. For rows 9, 11, and 13, the player acted after obtaining the object. The story ends when this eating action is performed three times. As with the action of eating, the system also ends the story if the player acts to obtain an object a certain number of times. Finally, rows 14 to 16 are ending sentences provided by the system. Once the system decides the story’s ending, it darkens the screen and outputs the ending.

3. Experiment

The purpose of this experiment was to evaluate the sentences generated by the system and its effectiveness. We thus evaluated the system’s fun, operability, and future usefulness. In addition, we used the semantic differential (SD) method and a questionnaire to evaluate the fun and naturalness of the generated stories.

We recruited 18 students majoring in computer science whose age was from 22 to 25 and gave them a reward after their participation in the experiment. The system described in this paper was developed and evaluated as a prototype to evaluate the use of action history with emotional expressions for review in learning. To familiarize themselves with the operation of the Oculus Touch,
Table 4. List of questionnaire items

<table>
<thead>
<tr>
<th>Questionnaire content</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1 The generated story is interesting.</td>
<td></td>
</tr>
<tr>
<td>Q2 The generated story is natural.</td>
<td></td>
</tr>
<tr>
<td>Q3 Between a story with emotion information and a story without it, which one gives you a better image of the situation or flow of the scene?</td>
<td>2-point scale</td>
</tr>
<tr>
<td>Q4 The tutorial is easy to understand.</td>
<td>5-point scale</td>
</tr>
<tr>
<td>Q5 The story generation phase is easy to operate.</td>
<td>5-point scale</td>
</tr>
<tr>
<td>Q6 I could easily make a story.</td>
<td>5-point scale</td>
</tr>
<tr>
<td>Q7 The story was made from my own actions.</td>
<td>5-point scale</td>
</tr>
<tr>
<td>Q8 I want to make a story using the system again.</td>
<td>5-point scale</td>
</tr>
<tr>
<td>Q9 I want to make a story in another situation (virtual space).</td>
<td>5-point scale</td>
</tr>
</tbody>
</table>

Table 5. List of scale of SD method

<table>
<thead>
<tr>
<th>Scale of SD Method</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural</td>
<td>Unnatural</td>
</tr>
<tr>
<td>Intellectual</td>
<td>Emotional</td>
</tr>
<tr>
<td>Funny</td>
<td>Boring</td>
</tr>
<tr>
<td>Various</td>
<td>Monotonous</td>
</tr>
<tr>
<td>Absurd</td>
<td>Sensible</td>
</tr>
<tr>
<td>Become bored</td>
<td>Concentrate</td>
</tr>
<tr>
<td>Dynamic</td>
<td>Static</td>
</tr>
<tr>
<td>Unique</td>
<td>Common</td>
</tr>
<tr>
<td>Active</td>
<td>Passive</td>
</tr>
<tr>
<td>Hard to read</td>
<td>Easy to read</td>
</tr>
</tbody>
</table>

Table 6. Questionnaire items and scoring

<table>
<thead>
<tr>
<th>Item</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly disagree</td>
<td>1.00</td>
</tr>
<tr>
<td>Disagree</td>
<td>2.00</td>
</tr>
<tr>
<td>Unsure</td>
<td>3.00</td>
</tr>
<tr>
<td>Agree</td>
<td>4.00</td>
</tr>
<tr>
<td>Strongly agree</td>
<td>5.00</td>
</tr>
</tbody>
</table>

Table 7. SD method items and scoring

<table>
<thead>
<tr>
<th>Item</th>
<th>Position on the questionnaire</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very negative</td>
<td>Left</td>
<td>-3.00</td>
</tr>
<tr>
<td>Rather negative</td>
<td>Left</td>
<td>-2.00</td>
</tr>
<tr>
<td>Fairly negative</td>
<td>Left</td>
<td>-1.00</td>
</tr>
<tr>
<td>Neither</td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>Fairly positive</td>
<td>Right</td>
<td>1.00</td>
</tr>
<tr>
<td>Rather positive</td>
<td>Right</td>
<td>2.00</td>
</tr>
<tr>
<td>Very positive</td>
<td>Right</td>
<td>3.00</td>
</tr>
</tbody>
</table>

the subjects first experienced a tutorial. The tutorial had simple objects such as a cube, and the subjects practiced performing all the operations in Table 3. To enable the subjects to enjoy the system’s content, we did not explain the magical tool in the tutorial. Instead, we let the subjects experience the tool directly in the system by explaining it to them orally and through the content.

For experimental material, we created a questionnaire to evaluate the stories generated by the system. The questionnaire was applied in terms of both the system and the generated narrative. Table 4 lists the items in the questionnaire. For each item, the subjects used a five-point scale consisting of “strongly disagree,” “disagree,” “unsure,” “agree,” and “strongly agree,” except for one item, Q3. In that case, Q3 was evaluated with three responses: “a story with emotion information,” “nothing,” and “a story without emotion information.” These questionnaire items also allowed free responses.

As additional experimental material to measure the subjects’ impressions of the whole generated story, we also created a questionnaire using the SD method. Table 5 lists the scales for the SD method and their positions on the questionnaire. The items of positive impression on each scale were described using net betting. The choices for the scale were “very negative,” “rather negative,” “fairly negative,” “neither,” “fairly positive,” “rather positive,” and “very positive.”

The following outlines the experimental method. The experiment time per subject was about 30 minutes.
1) Explanation of the experiment
2) System tutorial
3) Story generation by the system
4) Confirmation of the generated story
5) Questionnaire using the SD method
6) Questionnaire on the system and story

We first explained the details of the experimental procedure, including the system’s purpose. We then explained what kinds of operations could be done, by using both verbal and video instructions. After that, the subject was asked to use the system.

The flow of system use consisted of 5 minutes for the tutorial and 5 minutes for story generation. The time taken to generate a story changed the story’s flow depending on the actions taken in the virtual space. Therefore, the system was set to end the story when its branching condition was satisfied. After experiencing the system, the subject responded to the SD method questionnaire and the questionnaire on the system and the generated story. The purpose of the SD method questionnaire was to obtain the subjects’ intuitive impressions of the generated story, via a scale between pairs of adjectives. Note that we instructed the subjects to avoid responding with “neither” as much as possible, unless they could not determine the position between the two adjectives. The experiment ended once the questionnaires were complete.

As listed in Table 6, the questionnaire results were scored from 1 to 5 points for each item, ranging from “strongly disagree” to “strongly agree” for the items other than Q3. For tabulation of Q3, the response of “a story with emotion information” was counted as 1 point, while the other responses were each counted as 0. Then, the mean and standard deviation were calculated as descriptive statistics.

Then, as listed in Table 7, the SD method results from “very negative” on the left to “very positive” on the right were scored from -3 to 3 points, respectively, and then tabulated. For this scoring, if the mean value for an element of the SD method questionnaire was a negative number, it meant that the item on the left side was being evaluated. Similarly, a positive mean value meant that the item on the right was being evaluated. After tabulating the SD results, we calculated the mean and standard deviation as descriptive statistics and plotted a semantic profile. We also applied factor analysis by ProMax rotation of an unweighted least-squares method. All analyses used IBM SPSS Statistics 25.

### 4. Experimental Results

The responses to each questionnaire item were calculated as a percentage of the total number of responses. Table 8 summarizes these percentages. The discussion here is a simple comparison of descriptive statistics (percentages) only.

Table 9 lists the mean (\(M\)) and standard deviation (\(\text{SD}\)) for each scale of the SD method questionnaire. Figure 8 shows the semantic profiles for the mean values of each scale. In addition, Table 10 lists the factor analysis results, while Table 11 summarizes the factor correlation matrix. In Table 10, numerical values with factor loadings less than -0.4 or greater than 0.4 are shown in bold to make it easier to understand the factors having a strong correlation with the extracted factors. Finally, the factor analysis results were used to perform factor interpretation, as listed in Table 12. From these results, we could extract three factors, namely, non-immersivity, non-novelty, and narrativity.

| Table 8. Questionnaire results                                                                 |
|--------------------------------------------------|---------------------------------|-----------------|-----------------|-----------------|-----------------|
| Q1  | Story is interesting                      | Strongly agree | Agree | Unsure | Disagree | Strongly disagree |
| Q2  | Story is natural                          |                  |      |        |          |                  |
| Q4  | Tutorial is easy to understand            |                  |      |        |          |                  |
| Q5  | Easy to operate                           |                  |      |        |          |                  |
| Q6  | Easy to make story                        |                  |      |        |          |                  |
| Q7  | Story made from my own actions             |                  |      |        |          |                  |
| Q8  | Want to make story again                  |                  |      |        |          |                  |
| Q9  | Want to make story in another situation    |                  |      |        |          |                  |

| Table 9: | Questionnaire results                  |
|---------------------------------|---------------------------------|-----------------|-----------------|-----------------|-----------------|
| Q3  | Scene imagination                      | Sentence with emotion information | 94% | Sentence without emotion information | 6% |
Table 9. Descriptive statistics for each scale of the SD method

<table>
<thead>
<tr>
<th>Scale of SD method</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Became bored</td>
<td>1.22</td>
<td>1.83</td>
</tr>
<tr>
<td>Concentrated</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dynamic</th>
<th>Static</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dynamic</td>
<td>Static</td>
<td>0.922</td>
<td>-0.024</td>
</tr>
<tr>
<td>Natural</td>
<td>Unnatural</td>
<td>0.134</td>
<td>-0.542</td>
</tr>
</tbody>
</table>

Cumulative contribution rate (%)  

- Dynamic: 31.036
- Static: 46.445
- Unnatural: 56.880

**Figure 8. Semantic profiles.**

Table 12. Factor interpretation

<table>
<thead>
<tr>
<th></th>
<th>Factor</th>
<th>M</th>
<th>Narrativity</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Easy to read, Sensible</td>
<td>1.098</td>
<td></td>
</tr>
</tbody>
</table>
5. Discussion

In this research, we proposed a game-based system that generates stories including emotions and conducted an experimental evaluation of the system by subjects. The results showed that the system could relatively easily generate stories that interested the subjects. Among the free responses for Q7 in the questionnaire, “The story was made from my own actions,” subjects answered with “It was interesting that my actions were written and read objectively” and “It was fun to look back on my actions.” Therefore, we believe that generating action-based stories in a virtual space can appeal to users. Additionally, we found that the subjects enjoyed playing, because there was a significant difference in the questionnaire item Q8: “I want to make a story again.” Thus, players should be able to generate stories with various patterns by playing many times.

Unfortunately, the subjects were confused because the operation practice in the tutorial was not the same as the operation method in the virtual space. They needed to perform non-tool actions such as grabbing, throwing, and eating, but some subjects did not know how to perform those actions. Hence, the tutorial should allow the player to practice all the actions.

From the questionnaire, we found that it was easier for the subjects to imagine the scene by adding emotion information. The reason was that “because my emotions are recorded, it is easy to remember what I am doing.” There was also an opinion that “I hate insects, but I can picture throwing them later, so I can remember them.”

The “narrativity” factor obtained in the SD method questionnaire can be interpreted as showing that the generated narratives were consistent throughout. Therefore, the stories were considered to have a natural flow. The factor of “non-immersivity,” however, can be interpreted as indicating that the narrative generated by the player’s actions did not feel immersive. This suggests a need to improve the virtual space to allow more types of actions. Lastly, the “non-novelty” factor indicates that there was not much change in the scene or behavior. In other words, because the system’s virtual space was an average office, it may have lacked novelty. As there was a significant difference in the questionnaire item, “I want to make a story in another situation,” creating a new scene may lead to improvement in the system’s “novelty.”

The system records the history of the player’s actions and emotions, so if a player’s own game experience is interesting, it will be an interesting story for other people. If we think of this as a record of the learning experience in the game, if there is an impressive record with emotions, it will provide support for learning. The results suggest that historical information, in which a learner's actions and emotions are recorded in real time, is useful after the learner has experienced a game, because it enables the learner to remember his or her own experience. There is not much previous research on effective use of historical information in game-based learning. We believe, however, that it is effective to use the history of actions and emotions recorded in real time for learning.

In this paper, we have considered the effect of using action history with emotional expressions for review in learning. In this research, we investigated experiential learning content, but we have not yet conducted experiments on general learning support systems, so further investigation will be necessary. Furthermore, we need to investigate what kind of granularity and timing is effective in using emotional expressions for learning content.

6. Conclusion

In this paper, we introduced a game-based story generation system that automatically generates stories by using the player’s actions and emotions. Through experiments, we examined the usability, impression, and evaluation of the generated stories. The results suggest that historical information of a learner's actions and emotions recorded in real time is useful for the learner’s review after experiencing a game.
References


Booth, M. 2009. The AI Systems of Left 4 Dead. The ai systems of left 4 dead. Artificial Intelligence and Interactive Digital Entertainment Conference at Stanford.


León, C., & Gervás, P. (2014). Creativity in Story Generation From the Ground Up: Non-deterministic Simulation driven by Narrative. ICCC.


Investigating the Effects of Gamifying SQL-Tutor

Faiza TAHIRa*, Antonija MITROVICa & Valerie SOTARDIb

aIntelligent Computer Tutoring Group, Department of Computer Science and Software Engineering, University of Canterbury, New Zealand
bSchool of Educational Studies and Leadership, University of Canterbury, New Zealand
*faiza.tahir@pg.canterbury.ac.nz

Abstract: The practice of adding game elements to non-gaming educational environments has gained much popularity. Gamification has been found in some studies to increase learner engagement, motivation, and academic performance. However, there is a lack of empirical evidence to prove the effects of gamification in advanced learning technologies like Intelligent Tutoring Systems (ITS). This paper reports the results of an empirical study that included three categories of game elements (goals, assessment and challenges) implemented as badges in the context of SQL-Tutor. The study was conducted in a class under realistic conditions. SQL-Tutor was used voluntarily by 77 undergraduate students enrolled in a second-year database course. Although there were no differences between the experimental and control groups in terms of their interaction with SQL-Tutor and learning outcomes, we found a significant mediating effect of time-on-task on the direct relation between badges and achievement in the gamified condition. We also found evidence that not all students were interested in badges.

Keywords: Gamification, goals, assessment, challenges, badges, learning behavior, time-on-task, mediation effect.

1. Introduction

Engagement and motivation are crucial for effective learning. The amount of user interaction with an educational system is an important indicator of learning outcomes. In online learning, engagement refers to the student’s involvement with the system and motivation refers to his/her determination to achieve a goal. One strategy to increase motivation is gamification, i.e. the use of gaming elements such as leaderboards, points, badges and other virtual achievements common in games. These virtual achievements are not always connected to a tangible reward; they are meant to increase user involvement and their motivation to use those applications. For example, the TripAdvisor website (tripadvisor.com) rewards its users' points which do not have any monetary value. Badges are commonly used in educational environments. For example, PeerWise (Denny et al., 2018) awards virtual badges to students for writing or answering questions. Leaderboards are often used in applications where social activities are important, like comparing the performance of users in a course.

The term gamification was first used almost a decade ago (Deterding et al., 2011) and has gained much popularity. Gamification was found to be effective in many projects in maintaining user engagement by encouraging their actions and fostering quality and productivity of those actions (Hamari, 2013). However, the application of gamification in non-gaming environments does not always yield positive results. In a few cases, gamification may go unnoticed by users, and in other cases, it had negative effects on users which were completely unintended (Diefenbach & Müssig, 2019). Moreover, despite the growing number of educational environments incorporating gamification, there is a lack of empirical evidence proving its efficiency in a particular context/environment. Gamification might help in increasing engagement, enjoyment and motivation. However, if the learning environment is not proved to improve learning, gamification would not help. On the other hand, if an educational system is highly effective, gamification may not provide an additional benefit. Therefore, the process of applying gamification in a particular system should consider both the system’s effectiveness and the impact of gamification on the learner’s behavior.
Intelligent Tutoring Systems (ITSs) have a long history of proven results in education. There are many strategies used to address engagement and motivation in ITSs, such as supporting metacognitive strategies, e.g., self-regulation and self-assessment (Long & Aleven, 2013) and supporting affective states of learners. This study aims to explore the effects of gamification in SQL-Tutor (Mitrovic, 1998; Mitrovic, 2003), a mature ITS that teaches the Standard Query Language (SQL). The effectiveness of SQL-Tutor has been proven in multiple studies (Mitrovic & Ohlsson, 1999; Mitrovic, 2012). We start by providing a brief literature review of gamification and its effects. Section 3 presents our approach to gamifying SQL-Tutor, while Section 4 discusses the experiment design. We then present our findings in Section 5, and finally, the conclusion and limitations of the current work.

2. Related Work

Gamification is defined as “the use of game design elements in non-game contexts” (Deterding et al., 2011). It is considered to be less expensive in contrast to standalone games (Dicheva et al., 2015; Landers et al., 2017). As games are originally intended for enjoyment, gamification is also defined as motivational information systems which combine the efficiency of utilitarian systems and enjoyment of hedonic systems (Koivisto et al., 2019). Adoption of gamification is reported in many fields, particularly in education, health science and crowdsourcing. Several systematic literature reviews (Hamari, Koivisto, & Sarsa, 2014; Koivisto & Hamari, 2019) report that the most used game elements are points, badges and leaderboards, and the largest positive effects are on motivation and engagement, and less so on learning outcomes. However, not all studies report positive results, with some even reporting negative effects of gamification on students’ motivation and learning. Detailed analysis of these studies showed that they were focusing on behavioral changes of learners through the use of gamification and focused primarily on engagement, enjoyment and motivation. These reviews also point out methodological problems with the evaluations studies, which include small sample sizes, lack of control conditions, evaluating several gamification elements simultaneously and short duration of studies.

The theory of gamified learning proposed by Landers (Landers 2014; Landers et al., 2017) specifies that gamification has an effect on learning by influencing the learner’s behaviors or attitudes, via two theoretical paths. Some gamification elements influence learning behaviors/attitudes, which in turn directly influence learning outcomes; thus, the learning behavior acts as a mediator. In other situations, the influence of students’ behaviors or attitudes change the effectiveness of instructional content – that is, the learning behavior moderates the relationship between the content and learning outcomes. In a study using leader boards and the time-on-task as the mediating behavior, Landers and Landers (2014) found a significant improvement in learning.

Gamification has been applied to many web-based learning environments such as Code academy, Khan Academy and Stack Overflow (Marder, 2015; van Roy et al., 2018), and with mixed effects on student learning. Denny and colleagues (2018) conducted a study on Peerwise, a system for peer learning, with points and badges added as the gamification intervention and proved their effectiveness by targeting the engagement, motivation and self-testing behavior. In another similar study, gamification was examined on university students and computer games development course was gamified (O'Donovan, Gain, & Marais, 2013). The gamification elements were experience points, badges, leader boards, storyline and theme, presented with the help of gamified visuals. The study reported significant improvements in terms of student engagement and motivation, and the leader board was considered the biggest motivational element. The behaviors influenced most were attendance and attempting quizzes.

In another study, Haaranen and colleagues (2014) investigated the effects of badges in an online learning environment for a data structures and algorithm course. The badges were awarded for time management, early submissions and successfully completing exercises. The results showed that students were mostly indifferent about badges, and also the badges did not have significant effects on student behaviors and learning outcomes. The authors reported that students stopped working once they achieved enough scores for passing the course. However, no negative effects of badges were observed, and the authors suggested that the effects of gamification were highly context-dependent.

There is very little research focusing on gamification of ITSs. Long and Aleven (2014) explored the effects of two gamification features in Lynette (an ITS), which is re-practising of
previously completed problems and rewards for each completed problem. The results showed that
gamifying the ITS does not result in increased learning or enjoyment of students. However, the highest
learning gains were reported for those students who re-practised previously completed problems but
received no rewards on their performance (Long & Aleven, 2014). In the subsequent study (Long &
Aleven, 2016), Lynette rewarded students by awarding stars and badges when they selected unmastered
problems and showed perseverance on practising new problems. The gamification was shown to result
in higher learning outcomes compared to the control condition, as well as improved knowledge of the
problem-selection strategy.

This brief overview of literature acknowledges that three methodological gaps exist: 1) the
effects of gamification are highly context-dependent and may be overlooked in research designs, 2)
research on gamification inconsistently considers students’ behaviors or attitudes and 3) insufficient
design guidelines are available due to a lack of empirical studies. Our study attempts to fill these gaps.

3. Gamifying SQL-Tutor

We selected three categories of game elements from the nine categories discussed in the Theory of
gamified learning (Landers et al., 2017): goals, assessment and challenges. Challenges grow the
competition in students either in the form of standing in the class or achievement of the skill. Research
(Munshi et al., 2018) shows that student become bored/frustrated if they are not challenged enough.
Therefore, complex problems in the form of challenges can be helpful to retain their interest. Goals are
also considered as a form of challenge; however goal-setting theory states that goals can motivate
students if they are SMART (specific, measurable, achievable, realistic and time-bound) (Locke &
Latham, 1990; 2019). The goals selected in this study are according to these lines: they have only one
condition (specific), can be measured through completed problems (measurable), achievable, realistic,
and can be achieved within the 4-weeks study period (time-bound). The difference between challenges
and goals lies in the complex and hard to achieve challenges. SQL-Tutor provides assessments in the
form of pre/post-tests at the start/end of the study.

We implemented goals, assessment and challenges in SQL-Tutor via different types of badges
(Table 1). The goal-setting behavior is supported by fixing daily and weekly goals stated as winning
criteria for badges. The self-testing behavior is addressed by providing a quiz. Challenges are
implemented via several badges, and also as daily challenges, which consist of complex unsolved
problems. We hypothesize that all these game elements influence time-on-task, which has been shown
in many studies to influence learning outcomes (Landers et al., 2014; Denny et al., 2018).

<table>
<thead>
<tr>
<th>Group</th>
<th>Badge</th>
<th>Criterion</th>
<th>Behavior</th>
<th>Earned By</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>Go getter</td>
<td>Completing the first problem</td>
<td>Goal-setting</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>High flyer</td>
<td>3 problems in one session</td>
<td>Goal-setting</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>Achiever</td>
<td>5 problems in a day</td>
<td>Goal-setting</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>Activist</td>
<td>5 problems without complete solution</td>
<td>Challenge</td>
<td>16.66%</td>
</tr>
<tr>
<td></td>
<td>Leader</td>
<td>problem with the &quot;Group by&quot; clause</td>
<td>Challenge</td>
<td>16.66%</td>
</tr>
<tr>
<td>Classic</td>
<td>Energy house</td>
<td>6 problems in a row</td>
<td>Goal-setting</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>Scholar</td>
<td>5 problems/day for 5 consecutive days</td>
<td>Goal-setting</td>
<td>2.38%</td>
</tr>
<tr>
<td></td>
<td>Fireball</td>
<td>10 problems in one day</td>
<td>Goal-setting</td>
<td>92.80%</td>
</tr>
<tr>
<td></td>
<td>Champion</td>
<td>First daily challenge</td>
<td>Challenge</td>
<td>7%</td>
</tr>
<tr>
<td>Elite</td>
<td>Genius</td>
<td>Attempting the quiz</td>
<td>Self-testing</td>
<td>38.09%</td>
</tr>
<tr>
<td></td>
<td>Human dynamo</td>
<td>5 problems/day for 10 days</td>
<td>Goal-setting</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>Einstein</td>
<td>5 daily challenges over 2 weeks</td>
<td>Challenge</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>Live-Wire</td>
<td>5 problems per day for 20 days</td>
<td>Goal-setting</td>
<td>0%</td>
</tr>
</tbody>
</table>

The thirteen badges are divided into three groups: primary, classic and elite. The purpose of
primary badges is to grab the student’s attention at the early stage of using SQL-Tutor, such as awarding
a badge for solving the first problem, or for solving a problem using a difficult clause (group by). This
category also includes the Activist badge which discouraged the use of “complete solution”. Please note
that when the student submits a solution to SQL-Tutor, he/she can also specify the level of feedback.
The complete solution is the highest level of feedback in SQL-Tutor, which provides the full solution to the problem. Therefore, the Activist badge checks that the student solved the problem on his/her own, rather than copying the full solution provided by the system.

The classic group contains four badges, which emphasize practicing regularly, for example completing five problems for five consecutive days, and solving complex problems of the daily challenge. The last group, elite badges, consists of four badges and their main purpose is to keep engaging the student with SQL-Tutor over a longer period of time. In this category, badges are awarded when the student completes five problems every day for ten days, or solves five daily challenges in two weeks. The last badge awarded to those extraordinary students who completed five problems every day, for 20 consecutive days.

When the student fulfills the condition for a badge, he/she receives the notification about that badge immediately, as shown in Figure 1. Students can view all the badges awarded to them on the badge page, which also showed the badges which have not been achieved yet. SQL-Tutor also provides an Open Learner Model (OLM), in the form of skill meters. For the study, we modified the OLM page to show the next badge the student could achieve, as shown in Figure 2.
selected for a daily challenge need to be challenging for the student. SQL-Tutor summarizes the student’s learning progress using the student level, which ranges from 1 to 9. Problems in SQL-Tutor also have a complexity level (defined by the teacher) ranging over the same scale. Therefore, the problems selected for the daily challenge are previously unsolved problems, which satisfy two conditions: 1) their level of complexity is equal to the current student level or one level higher, and (2) these problems require the clauses of the SELECT statement which the student needs to practice (as per the student model). Each day, the daily challenge is presented to the student upon logging in, and is also available on the problem-selection page. Two badges (Champion and Einstein) are awarded when the student completes the first daily challenge, or when the student completes five daily challenges over two weeks respectively.

We also developed a quiz, consisting of seven multiple-choice questions and two true/false questions. The Genius badge is awarded for attempting the quiz, independently on the score achieved. When the student completes a quiz, the scores is shown immediately, so that the student can reflect on his/her knowledge. Awarding badge on attempting the quiz maximizes the effects of students’ self-testing abilities.

4. Experimental Procedure and Hypotheses

The participants were recruited from the 198 students enrolled in the second-year course on relational database systems at the University of Canterbury in 2019. Before the study, the students have learnt about the relational data model and SQL in lectures and had two labs sessions, in which they created tables and performed basic SQL queries in Oracle. The students were introduced to SQL-Tutor in a lab session. The use of SQL-Tutor was voluntary; the students did not receive any course credit for solving problems in SQL-Tutor. All enrolled students were randomly allocated to the control group (using the standard version of SQL-Tutor) or the experimental group, who used the gamified version. We obtained informed consent from 77 students (25% female, 62% male, 13% not specified); 42 in the experimental group and 35 in the control group.

The study lasted for four weeks. When students logged into SQL-Tutor for the first time, they received the pre-test, a short demographic questionnaire and a question about their previous experience of using gamification. The students could use SQL-Tutor whenever they wanted. The quiz was given at the end of the second week of the study to both control and experimental groups. The pre/post-test and the quiz were of similar complexity; each contained seven multiple-choice questions and two true/false questions (worth one mark each).

The post-test was administered online at the end of the fourth week. A major piece of the course assessment was the lab test focusing on SQL, worth 20% of the final grade. The lab test was given two days after the post-test. After the lab test, the students were invited to complete a survey. There were two versions of the survey. For the experimental group, there were four questions related to their opinion of the badges, and two questions related to daily challenges. Both groups received two questions about the quiz. The responses to these questions were recorded on the 5-point Likert scale, from ‘strongly disagree’ (1) to ‘strongly agree’ (5).

We made the following hypotheses, based on the results from literature (e.g. Landers & Landers, 2014), and from our own experience:

**H1**: The time-on-task is positively correlated with learning outcomes.

**H2**: The experimental group participants will spend more time solving problems in SQL-Tutor in comparison to the control group.

**H3**: Badges will have a mediating effect on learning outcomes, by influencing the time-on-task.

5. Results

The average score on the pre-test was 58.73% (sd = 26.05). The students interacted with SQL-Tutor on 3.39 days (referred to as Active Days) over four weeks (sd = 2.69, min = 1, max = 12), spending 260 min (min = 41, max = 1,441, sd = 243) in the system. During that time, the students solved an average of 37.47 problems (sd = 34.74, min = 3, max = 204). Only 28 students completed the post-test; we believe the reason for the low completion rate was that the post-test was not mandatory. In addition, the
post-test was given to the students only two days before the lab test. The average score on the post-test was 69.05\% (ds = 25.90). For the lab test, the average score was 60.83\% (sd = 17.07). In addition to defining queries, which students practiced in SQL-Tutor, the lab test covered other SQL topics, and therefore the lab test cannot be considered as the direct learning outcome. For those reasons, we use the student level at the end of the interaction with SQL-Tutor as a measure of students’ learning. The average student level was 3.56 (sd = 1.66, min = 1, max = 8). In the experimental group, 66\% of students reported having used some form of gamification, compared to 57\% of the control group participants.

5.1. Evaluating the Hypotheses

To evaluate H1, we regressed the student level on time-on-task. The time-on-task strongly predicts the student level ($\beta = .536$), and was statistically significant ($t = 5.5, p < .001$). Variance in student level explained by time-on-task was 28.7\%. Therefore, hypothesis H1 was supported.

Table 2 presents statistics for the two groups. There was no significant difference on the pre-test scores of the two groups, showing that the students had comparable levels of pre-existing knowledge. The experimental group students spent more time on task, had more sessions, attempted and solved more problems, and attempted more complex problems in SQL-Tutor in comparison to the control group, although none of the differences are significant. Therefore, our hypothesis H2 is not supported. There was also no significant difference between the groups on the number of active days, student levels, the post-test and lab test scores.

Table 2. Summary statistics of SQL-Tutor usage: mean (sd)

<table>
<thead>
<tr>
<th></th>
<th>Experimental (42)</th>
<th>Control (35)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test %</td>
<td>59.52 (24.02)</td>
<td>57.78 (28.62)</td>
</tr>
<tr>
<td>Time-on-task (min)</td>
<td>288.40 (302.02)</td>
<td>225.94 (143.44)</td>
</tr>
<tr>
<td>Sessions</td>
<td>7.29 (7.84)</td>
<td>6.11 (4.49)</td>
</tr>
<tr>
<td>Active Days</td>
<td>3.33 (3.09)</td>
<td>3.46 (2.13)</td>
</tr>
<tr>
<td>Attempted problems</td>
<td>42.26 (42.75)</td>
<td>37.34 (26.94)</td>
</tr>
<tr>
<td>Solved Problems</td>
<td>39.33 (40.99)</td>
<td>35.23 (25.72)</td>
</tr>
<tr>
<td>Max Problem Complexity</td>
<td>6.95 (1.78)</td>
<td>6.71 (2.02)</td>
</tr>
<tr>
<td>Student level</td>
<td>3.31 (1.62)</td>
<td>3.86 (1.68)</td>
</tr>
<tr>
<td>Post-test %</td>
<td>n = 17, 67.97 (26.32)</td>
<td>n = 11, 70.71 (26.42)</td>
</tr>
<tr>
<td>Lab test %</td>
<td>60.43 (16.49)</td>
<td>61.31 (17.97)</td>
</tr>
</tbody>
</table>

To evaluate H3, we used the data for the experimental group only. We analyzed the mediation effect using the Process macro, version 3.5 software for SPSS (Hayes, 2017), with the student level as the dependent variable. Figure 3 shows the standardized regression coefficients for the mediation model. The direct effect of badges on the student level is not significant ($p = .08$), but the significant relationship in this first step is not a requirement for mediation (Shrout & Bolger, 2002). The direct effect of badges on time is significant ($p < .001$), as is the direct effect of time on the student level ($p$,

![Figure 3. The mediation model, with standardized coefficients](image)
The indirect and total effects in the model are tested using bootstrap samples and 95% confidence intervals. Results show that the standardized, indirect effect of badges on the student level is $\beta = 0.32$. The confidence interval for the estimate of the indirect effect $[.165, .501]$ does not include zero; therefore the null hypothesis is rejected. 52.26% of the total effect is mediated. The Sobel test of significance of mediation gives 2.62 ($p < .01$), indicating that time on task mediates the direct relationship between the number of badges and the student level. Therefore, hypothesis H3 is confirmed.

5.2. Further Investigation of the Experimental Group

Overall, the experimental group students achieved from 4 to 7 badges, with a mean of 5.43 (sd = .86). The percentage of students from the experimental group who earned various badges is shown in the last column of Table 1. On the very first day of interacting with SQL-Tutor, the students achieved an average of 4.60 badges (sd = .76). Only seven students achieved all primary badges; therefore they were the only ones who were given daily challenges. For that reason, it is not possible to make any conclusions about the daily challenges.

The literature review shows that in some cases, students are not interested in badges when they are not directly related to course credit. To investigate whether there is a difference in how much the experimental group students were interested in badges, we divided the experimental group students into two subgroups: those who visited the badge page at least once (23 students), and those who have never visited that page (19 students). Table 3 presents the differences found between the two subgroups.

Table 3. Comparing experimental group students who visited the badge page or not: mean (sd)

<table>
<thead>
<tr>
<th></th>
<th>Seen badge page (23)</th>
<th>Not seen (19)</th>
<th>Significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test %</td>
<td>54.59 (25.05)</td>
<td>65.49 (21.88)</td>
<td>$p = .22$</td>
</tr>
<tr>
<td>Time-on-task (min)</td>
<td>365.30 (272.27)</td>
<td>195.32 (316.96)</td>
<td>$U = 348.5, p &lt; .001$</td>
</tr>
<tr>
<td>Sessions</td>
<td>9.48 (7.69)</td>
<td>4.63 (7.37)</td>
<td>$U = 334.5, p &lt; .005$</td>
</tr>
<tr>
<td>Active Days</td>
<td>4.13 (3.22)</td>
<td>2.37 (2.71)</td>
<td>$U = 312.5, p &lt; .05$</td>
</tr>
<tr>
<td>Attempted problems</td>
<td>51.91 (39.51)</td>
<td>30.58 (44.62)</td>
<td>$U = 332, p &lt; .005$</td>
</tr>
<tr>
<td>Solved Problems</td>
<td>47.48 (36.86)</td>
<td>29.47 (44.49)</td>
<td>$U = 326.5, p &lt; .01$</td>
</tr>
<tr>
<td>Constraints</td>
<td>287.74 (60.98)</td>
<td>247.84 (75.82)</td>
<td>$U = 299.5, p &lt; .05$</td>
</tr>
<tr>
<td>Badges</td>
<td>5.74 (.81)</td>
<td>5.05 (.78)</td>
<td>$U = 317, p &lt; .01$</td>
</tr>
<tr>
<td>Student level</td>
<td>3.70 (1.72)</td>
<td>2.84 (1.39)</td>
<td>$p = .07$</td>
</tr>
<tr>
<td>Post-test %</td>
<td>n = 13; 4.38 (2.93)</td>
<td>n = 8; 5.88 (3.72)</td>
<td>$p = .34$</td>
</tr>
<tr>
<td>Lab test %</td>
<td>59.74 (13.90)</td>
<td>61.26 (19.55)</td>
<td>$p = .81$</td>
</tr>
</tbody>
</table>

There was no significant difference between the two subgroups on the pre-test scores. The students who visited the badge page have interacted with SQL-Tutor significantly more, measured either as the total time ($p < .001$), the number of sessions ($p < .005$), or the number of active days ($p < .05$). Those students attempted/solved more problems ($p < .005$ and $p < .01$ respectively) than their peers, and also achieved significantly more badges ($p < .01$). The students who have seen more badges have used significantly more constraints than their peers. In SQL-Tutor, domain knowledge is represented in terms of more than 700 constraints. Therefore, the students who visited the badge page covered a higher proportion of the domain in comparison to their peers. Therefore, there is evidence that visiting the badge page is correlated with more time-on-task and engagement. However, there was no significant difference between the two subgroups in terms of learning, measured either by the student level achieved ($p = .07$), post-test scores ($p = .34$) or the lab test score ($p = .81$).

5.3. Self-testing Behavior

As mentioned in Section 4, the quiz was completely optional and provided to both experimental and control groups. To analyze students’ self-testing behavior, we investigated whether there is a difference in the student level achieved based on whether the students took the quiz and the group they were in (Table 4). We introduced a dummy QuizTaken variable, with values of 0 (quiz not taken) or 1 (quiz
taken). In the control group, 12 students attempted the quiz while 23 did not. For the experimental group, 14 out of 42 students attempted the quiz. A two-way ANOVA ($F = 3.07, p < .05$, partial $\eta^2 = .11$) revealed neither a significant interaction between group and QuizTaken, nor the main effect of group, but there was a significant effect of the self-testing behavior ($p = .01$, partial $\eta^2 = .09$) Students who attempted the quiz achieved a significantly higher student level.

<table>
<thead>
<tr>
<th>Table 4. Student level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Control</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Exper.</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Table 5 presents the statistics for students who attempted or did not attempt the quiz. There was no significant difference on the pre-/post-test scores and the lab test scores. The students who attempted the quiz interacted with SQL-Tutor significantly more, measured in terms of time, sessions, active days and attempted/solved problems. They used more constraints and solved more complex problems, thus achieving higher student levels.

<table>
<thead>
<tr>
<th>Table 5. Comparing students who attempted/not attempted the quiz: mean (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not attempted (51)</td>
</tr>
<tr>
<td>Pre-test %</td>
</tr>
<tr>
<td>Time-on-task (min)</td>
</tr>
<tr>
<td>Sessions</td>
</tr>
<tr>
<td>Active Days</td>
</tr>
<tr>
<td>Attempted problems</td>
</tr>
<tr>
<td>Solved Problems</td>
</tr>
<tr>
<td>Max Problem Complexity</td>
</tr>
<tr>
<td>Constraints</td>
</tr>
<tr>
<td>Student level</td>
</tr>
<tr>
<td>Post-test %</td>
</tr>
<tr>
<td>Lab test %</td>
</tr>
</tbody>
</table>

5.3 Survey Responses

We received 21 survey responses from the experimental group and 22 responses from the control group students. Table 6 summarizes the responses to the four questions on badges from the experimental group students. The Cronbach alpha for those questions is 0.88.

<table>
<thead>
<tr>
<th>Table 6. Responses from the experimental group (1 - strongly disagree to 5 - strongly agree)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question</td>
</tr>
<tr>
<td>Badges motivated me to participate more than I would have otherwise.</td>
</tr>
<tr>
<td>I found being able to earn badges increased my enjoyment of using SQL-Tutor</td>
</tr>
<tr>
<td>I would prefer not to see badges in SQL-Tutor.</td>
</tr>
<tr>
<td>The badges awarded for solving problems motivated me to solve more problems than I would have otherwise.</td>
</tr>
</tbody>
</table>

The responses of the experimental group indicate that students did not find badges very motivating. Students were indifferent in their responses about the enjoyment when they received badges. However, 39% of students stated they wanted to see the badges. We do not discuss the questions on daily challenges, as only seven students received them during the study. Almost 62% of students wanted to see the daily challenges in SQL-Tutor; this figure reveals that students were interested in daily challenges in principle. The students from both groups enjoyed attempting quiz
6. Conclusions

This paper presents a classroom study in which we analyzed the effect of gamification in the context of SQL-Tutor. Our findings highlight the effects of gamification in the context of an ITS, under realistic conditions, in a study that lasted four weeks.

Starting from Lander’s theory of gamified learning (2014), we designed badges which supported goal setting, assessment and challenges—three common categories of game elements. We hypothesized that the badges would motivate students to spend more time on task (i.e. problem solving in SQL-Tutor). The goal-setting behavior is supported by setting SMART goals/criteria for achieving each badge. Challenges motivate students to perform more complex tasks, and the quiz allowed students to test their knowledge.

Our study provides initial evidence that badges can positively increase student achievement in ITSs (measured as the student level achieved in SQL-Tutor), and that this relation can be mediated by the amount of time participants spend on the task. The results show the impact of gamification on learning through behavioral change, supporting the theory of gamified learning with the time-on-task as a valid behavior target for gamification. From the statistical analysis, we first determined that time-on-task correlates and predicts learning outcomes. We did not find a difference between gamified and non-gamified groups in terms of time spent in SQL-Tutor, problems completed, and learning outcomes. A possible explanation for this finding is that the students are already highly motivated, and used SQL-Tutor to prepare for the lab test. However, we found evidence that goal-setting, challenges and self-testing behaviors implemented as badges indirectly and significantly affect learning outcomes through the time-on-task as the mediator.

There are two major limitations of our study, the first being the small sample size. The second limitation was the design of the badges, which could be designed in a more visually attractive manner. As discussed, almost 46% of students in the experimental group did not access the badge page despite receiving badge notifications. This shows that the design of badges was not attractive enough to entice some learners and motivate them to achieve.

7. Acknowledgements

We thank Jay Holland for helping with the study, as well as our participants.

8. References


Design of a Game-based Intelligent Learning Environment to Remediate Fraction Addition/Subtraction Misconceptions

Jarl Brent OBEDOZA* & Raymund SISONb*

a,b College of Computer Studies, De La Salle University, Philippines
b College of Education, De La Salle University, Philippines
*jarl_brent_obedoza@dlsu.edu.ph, raymund.sison@delasalle.ph

Abstract: Fraction addition and subtraction entail arithmetic procedures that are difficult for elementary students. Difficulties come from the inherent complexities of these procedures. A proposed method to deal with these complexities involves a game-based intelligent learning environment (GILE), the learning outcome mechanics and other game elements of which are designed based on the literature on fraction instruction and fraction misconceptions. Preliminary results seem to point toward the ability of GILE to improve learner performance and remediate misconceptions for gamers and those whose procedures are close to the mechanics.

Keywords: Game-based Intelligent Learning Environment, Educational Games for Fraction Addition/Subtraction, Remediating Fraction Addition/Subtraction Misconceptions

1. Introduction

Fraction arithmetic is difficult for elementary students. Difficulties they experience hinder learning of the proper procedures and concepts to perform fraction arithmetic (Lortie-Forgues et al., 2015). Because of these difficulties, misconceptions arise when students perform either fraction addition or subtraction, such as adding the numerators and denominators of 2 fractions separately (e.g. 1/2 + 2/3 = 3/5) (Aksoy & Yazlik, 2017; Siegler et al., 2011; Fazio & Siegler, 2011; Mohyuddin & Khalil, 2016). For this reason, teachers, researchers, and other concerned parties implement various techniques to improve students’ learning of fraction arithmetic.

Some of the techniques involved are those used in the classroom environment. Some of these include the use of visuals, which make it easier for students to understand what is taught (Lamon, 2012; Nardi, 2014). An example of this is the use of bar models in Singapore math (Hoven & Garelick, 2007), where fractions are represented as concrete objects or pictures with associated numbers first before they are shown as abstract symbols. Physical manipulatives (Gabriel et al., 2012) take it a step further by allowing interaction. These help the students understand the mechanisms of fraction arithmetic in a concrete, tangible manner.

However, traditional classroom instruction has its limitations, such as the ratio of the teacher to the students (Schanzenbach, 2014). Teachers would have to cater to a group instead of the individual needs of students, causing inefficiencies when dealing with student misconceptions.

This issue can be addressed with the help of technology, such as intelligent tutoring systems (ITSs). ITSs solve this issue by providing students with individualized learning. They can guide students whenever they perform an error during the arithmetic process, allowing them to correct themselves (e.g. AnimalWatch, Beal, 2013). Systems also exist that make students commit errors (e.g. through trick questions) to diagnose problems for immediate correction (Layton, 2016). ITSs dealing with fraction arithmetic have been developed with good results (e.g., Beal, 2013; Riconscente, 2013). However, their explanations of fraction addition/subtraction procedures, if any, are through abstract symbols.

While ITSs can simulate tutors, the current generation of children learn differently than those of older generations (Prensky, 2001b). This indicates the need for a different approach, especially one that
takes into consideration that which is familiar to the current generation, what they experience, and how they behave or think. Serious games, which are video games designed for productivity, provide individualized learning as well, and use engagement to motivate learning in an interactive learning environment, allowing the student to actively participate in the learning process, similar to how physical manipulatives are used.

A few serious games have been developed for fraction arithmetic that use visual models. Slice Fractions (Cyr et al., 2016) has been shown to enable students to learn subtraction of similar fractions, but not of fractions with different denominators, which necessitates the students’ finding a common denominator. Discord (Espulgar et al., 2018) covers addition and subtraction of dissimilar fractions, but its mechanics might be too complex for some learners.

For students to overcome the difficulties of fraction addition and subtraction, this research uses a game-based intelligent learning environment, a medium familiar to their generation, to implement an intuitive, comprehensible approach to alleviate difficulties by addressing misconceptions.

2. Game-based Intelligent Learning Environments for Fraction Addition/Subtraction

2.1 Game-based Intelligent Learning Environments

An intelligent learning environment (ILE) is a computer program that adaptively provides a set of artifacts or tools that a learner can manipulate so that in the end, he or she will have learned a target concept or skill (Sison, 2001). To adapt to what the learner is currently doing or having difficulty with, an ILE uses a student model, also known as a learner model. A student model is a possibly partial and certainly approximate representation of the learner’s knowledge, including misconceptions, as well as the learner’s goals, preferences, and idiosyncrasies (Sison and Shimura, 1998; Woolf, 2010, p. 48). An intelligent tutoring system (ITS) is another kind of program for computer-assisted learning that makes use of a student model. It uses a student model to diagnose and remediate its learners. However, whereas an ITS performs the remediation explicitly, an ILE does this more subtly.

A game-based intelligent learning environment (GILE) is an ILE that is in the form of a videogame. It therefore falls under the category of a serious game or an educational game. However, it has the additional capability of understanding a user’s misconceptions, if any, and can attempt to enable the user to recognize and correct these. Therefore, a GILE has three components: a game component, a learning modeling component, and a pedagogical component (Figure 1).

![GILE Components and Interactions](Image)

Figure 1. GILE Components and Interactions.
Fraction addition and subtraction is difficult to learn, and Lortie-Forgues et al. (2015) have identified seven sources of difficulty found in fractions and decimal arithmetic, including the following.

1. Fraction Notation - Due to how a fraction is represented (in the form of a/b, b is not equal to 0), some students get confused and treat the numbers as independent numbers, as an arithmetic operation (e.g. a+b), or a single number (e.g. a/b is written as ab). This also increases cognitive load on the student when performing arithmetic operations.

2. Accessibility of Fraction Magnitudes – Understanding the magnitude of fractions is complex compared to that of whole numbers. Its factors (numerator and denominator) must be derived to get the representative magnitude, whereas whole number magnitudes can be understood as is. Knowledge of this has been statistically tested to be related to success in fraction arithmetic (Siegler et al., 2011).

3. Opaqueness of Standard Fraction Arithmetic Procedures – It is not apparent to some students why procedures during fraction arithmetic must be done. For example, why is there a need to make denominators equal before addition or subtraction can proceed? While there is an explanation, Lortie-Forgues et al. (2015) considers it advanced for the grade level where fractions are being taught (i.e. requiring algebra to explain).

4. Complex Relations between Rational and Whole Numbers Arithmetic Procedures - When adding or subtracting fractions, only the numerators are added or subtracted as if they were whole numbers but not the denominators, making the procedure confusing for students.

5. Sheer Number of Distinct Procedures - To perform addition or subtraction of fractions, one must ensure that the denominators are equal, making multiplication a necessary step, before the actual operation is done. This step requires mastery of fraction equivalence, which is a separate procedure than fraction addition or subtraction.

There are two other sources of difficulties, but they do not directly relate to fraction addition and subtraction.

These difficulties would explain why students commit errors. For example, they would respectively add the numerators and denominators of two fractions, an error associated with the misconception of numerator and denominator being two different independent natural numbers (Aksoy & Yazlik, 2017; Siegler et al., 2011; Fazio & Siegler, 2011; Mohyuddin & Khalil, 2016). A possible explanation for this is what Siegler et al. (2011) call the whole number bias, which is the knowledge of whole numbers interfering with the learning of fractions (Lamon, 2012). That is, students carry over knowledge from previous experiences and apply it to an inappropriate context (Layton, 2016), which points to the sources of difficulty identified (numbers 1 and 4 in the list above).

Siegler et al. (2011) report that adding numerators and denominators are more frequent in fraction addition/subtraction involving different denominators. This points to the difficulty students face when trying to make dissimilar fractions become similar. This is the procedure in the operation where fractions of different denominators are made to have similar denominators. For example, Idris & Narayanan (2011) and Abdul Ghani & Maat (2018) identified an unusual error where students would add numerators and choose one denominator as the denominator sum. Aksoy & Yazlik (2017) describes it as either not knowing how to perform denominator equalization or being unfamiliar with the operation. This points to the third and fifth difficulties in the list above.

Considering the situations stated above, a way to address the difficulty in equalizing the denominators of fractions is necessary. To do so, we must understand the operations to see why and how dissimilar fractions are made similar.

2.3 Bar Models in Singapore Math

Singapore math is based on James Bruner’s theory that people learn in 3 steps: concrete, iconic, and abstract. They start by learning concepts through common objects, then associate these objects with abstract representation, before finally working with purely abstract representations. Bar models can be used in this manner by letting students work with visual objects to use procedures and operations on (e.g. addition can be taught by putting together bars of the same sizes). This way, students can focus on the underlying mechanisms of arithmetic operations without worrying about abstract notations. Abstract
representations can then be gradually introduced, until the bar model can be removed. Slice Fractions (Cyr et al., 2016), a game for teaching fraction subtraction, has been shown to successfully use bar models through its game mechanics. However, students do not find common denominators of fractions by themselves. Discord (Espulgar et al., 2018) covers addition and subtraction of dissimilar fractions, but its mechanics might be too complex for some learners.

3. System Design and Implementation

Endless Sky is the name of the educational mobile game presented in this paper. It is a game-based intelligent learning environment (GILE) for fraction addition and subtraction, developed in Unity 2019.1.5f1.

3.1 Outcome-based Game Design

The design process of Endless Sky follows the iterative and incremental outcome-based game design methodology of Sison et al. (2018), which begins with, and is based on, intended learning outcomes (LOs). A learning outcome (LO) is a desired, tangible capability that a student can demonstrate after a learning experience (Spady, 1994). In an outcome-based educational approach, what the student learns at the end of the lesson determines what the curriculum and/or pedagogy is (Spady, 1994).

After the LOs have been identified, a suitable game genre and premise are then designed based on the LOs, after which special game mechanics, called LO mechanics are crafted.

To be effective, LO mechanics and the other formal and dramatic elements of a GILE must be designed with the help of established guidelines (e.g., Sweetser et al., 2017) as well as what the domain’s literature suggests to be effective ways of teaching the domain, as well effective ways of addressing the major difficulties and misconceptions of learners in the domain, as discussed in Sections 2.2-2.4 above.

3.2 Learning Outcomes

Endless Sky focuses on three learning outcomes regarding fraction addition and subtraction. These three are specified in the Department of Education’s (DepEd’s) Basic K to 12 Curriculum (Tabilang et al., 2015) and are listed in Table 1.

<table>
<thead>
<tr>
<th>Learning Outcome</th>
<th>Associated DepEd Learner Competency Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correctly solve fraction addition and subtraction problems with similar denominators</td>
<td>M4NS-IIg-83</td>
</tr>
<tr>
<td>Transform fractions with different denominators into fractions with similar denominators</td>
<td>M4NS-IIc-69.1</td>
</tr>
<tr>
<td>Correctly solve fraction addition and subtraction problems that have different denominators</td>
<td>M4NS-IIg-83</td>
</tr>
</tbody>
</table>

3.3 Game Genre and Premise

Endless Sky is an endless puzzle game. It is non-time-bound, so players can take their time solving the fraction addition and subtraction puzzles. The fraction addition and subtraction puzzles are randomly generated and become progressively challenging as the player progresses in the game. As it is an endless game, the player can continue playing until they trigger the game-over or feel like ending the session.

The game is set in a world with a vertically boundless sky that has, unfortunately, begun to collapse. Rifts in the sky have begun to form, causing sky fragments to fall. The player is tasked to prevent the sky from collapsing by sealing the rifts using the fallen sky fragments, ascending higher into the now distorted, endless sky.
The endless sky is actually made of a material that is made up of cells, which in turn can have any number of partitions. When a rift occurs, a fragment (i.e., a set of adjacent partitions) of a cell breaks off and falls to the earth. Each rift can therefore be viewed as a fraction and is represented visually using a bar model.

3.4 LO Mechanics and Other Game Elements

To help students achieve the LOs shown in Table 1, Endless Sky uses four LO mechanics, shown in Table 2.

<table>
<thead>
<tr>
<th>LO</th>
<th>LO Mechanic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correctly solve fraction addition and subtraction problems with similar denominators</td>
<td>Merge (for addition)</td>
<td>Combine two or more sky fragments with similar partitions.</td>
</tr>
<tr>
<td>Cleave (for subtraction)</td>
<td>Use one sky fragment to detach pieces from another sky fragment with similar total pieces.</td>
<td></td>
</tr>
<tr>
<td>Split (for subtraction)</td>
<td>Split a sky fragment into two smaller sky fragments.</td>
<td></td>
</tr>
<tr>
<td>Transform fractions with different denominators into fractions with similar denominators</td>
<td>Scale</td>
<td>Change the total pieces of a sky fragment based on a multiplier.</td>
</tr>
<tr>
<td>Correctly solve fraction addition and subtraction problems that have different denominators</td>
<td>Scale+Merge</td>
<td></td>
</tr>
<tr>
<td>Scale+Cleave</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scale+Split</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The objects in the game that the user interacts with are the sky fragments that have fallen from the sky. Like rifts, fragments are fractions, and are also represented visually as bar models. To seal a rift, one can only use a fragment of the same size as the rift. So, a rift of 1/3 (of a sky cell) can only be sealed by a fragment with a value of 1/3 (of a sky cell). Attempting to seal a rift with the wrong-sized fragment will only create more rifts in the sky (as well as decrease the player’s score). However, as Table 2 and Figure 2 show, it is possible to merge two sky fragments, split a sky fragment into two, cleave two sky fragments (detaching pieces of a sky fragment based on the number of pieces of another), or scale a sky fragment (not by changing its size but by changing the number of its partitions).

Figure 2. Tutorial for merging 2 or more sky fragments (top left), cleaving 2 sky fragments (top right), splitting a sky fragment into smaller ones (bottom left), and scaling sky fragments to adjust their total pieces (bottom right).

Figure 3 shows how the dissimilar fractions could be added. This corresponds to the third LO in Table 2, which involves two LO mechanics: scaling and either merging, cleaving, or splitting. In the
figure, the rift 2/3 needs to be sealed, using the fragments 1/2 and 1/3. First, 1/2 is scaled to 2/4 (Figure 3-3), then 1/3 is scaled to 2/6 (Figure 3-4), then 2/4 is scaled to 3/6 (Figure 3-5), and the two are merged into 5/6 (Figure 3-6). Finally, 5/6 is split into 4/6 and 1/6 (Figure 3-7), in preparation for the scaling down of 4/6 to 2/3 and the eventual sealing of the rift 2/3. It should be noted that this is a more detailed step to show how the arithmetic process is adapted into game mechanics. Players can skip certain steps (such as 3-3) if they know, for example, that they should scale both fragments to have 6 total partitions. The game also allows, for example, the rift 2/3 with the fragment 4/6 to show fraction equivalence in case the player is knowledgeable about that concept.

![Figure 3. Scaling 2 dissimilar fractions represented as sky fragments in the game into similar fractions so they can be Merged into a new sky fragment that can be used to fill a rift the given blocks cannot fill.](image)

As mentioned earlier, attempting to seal a rift with a fragment of the wrong size will only create more rifts. Moreover, a player can only incur one unsuccessful seal attempt per sky level. The second unsuccessful seal attempt will cause the sky level to collapse, bringing the sky, as well as the game, to an end. Of course, the student can always play again.

Sealing the rifts at a given sky level will enable the player to move up to the next sky level. The number of sky rifts that need to be sealed and the number of sky fragments that the user can work with are computed based on the player’s performance, which is monitored by the learner modeling component, which will be discussed next.

Following the outcome-based design methodology, the difficulties that were enumerated in section 2.2 were also taken into consideration in the design of the other formal and dramatic elements of the game. Table 3 describes how these difficulties were addressed in Endless Sky.

Table 3. Fraction Learning Difficulties and In-Game Elements that Address Them.

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>How the Game Addresses the Difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fraction Notation</td>
<td>The game transitions from showing pure visual representation of the fractions through the sky fragments, to visuals accompanied with the usual fraction notation, to only the fraction notation to allow the player to gradually understand fraction notation and see how they are visually represented. Through the LO mechanics (e.g. splitting, merging), the player is shown how the fraction notation and the visuals interact with respect to their actions (e.g. increasing or decreasing the numerator value, combining a group of sky fragments together).</td>
</tr>
<tr>
<td>Accessibility of Fraction Magnitudes</td>
<td>Visualization and hints using animation show players why groups of sky fragments cannot be merged/cleaved through emphasis on the differing number and sizes of the sky fragments’ pieces and/or their denominator values.</td>
</tr>
<tr>
<td>Complex Relations between Rational and Whole Numbers Arithmetic Procedures</td>
<td>The game lays down a simple rule that they cannot merge/cleave sky fragments with differencing pieces and/or denominators and giving them freedom to work out the problem however they like. Being a video game, the LO mechanics streamline the arithmetic procedure, so the player need only worry about what steps to do to solve a problem, rather than worry about non-essentials.</td>
</tr>
</tbody>
</table>
3.5 Learner Modeling Component

The learner model of Endless Sky is implemented using a Bayesian network (BN), which the learner modeling component manages. Figure 4 shows the BN’s structure. Updates occur every time a puzzle is cleared, i.e., all rifts are sealed. The BN implementation used in Endless Sky reuses the probabilistic computation library used in (Madrigal et al., 2018) and Espulgar et al. (2018).

Figure 4. BN representing the Learner Model, with arrows indicating probability propagation direction.

In addition to the BN, Endless Sky also uses a table of misconceptions (Table 4). These misconceptions are based on Aksoy & Yazlik (2017), Siegler et al. (2011), Fazio & Siegler (2011), and Mohyuddin & Khalil (2016). When a player makes an unsuccessful attempt to merge/cleave fragments, Endless Sky tries to map the error into one of the two misconceptions in Table 4.

Table 4. Misconceptions Table

<table>
<thead>
<tr>
<th>Misconception</th>
<th>Error Variant</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numerator and denominator are 2 independent natural numbers</td>
<td>Adding numerators and denominators independently</td>
<td>2/3 + 1/3 = 3/6</td>
</tr>
<tr>
<td></td>
<td>Subtract numerators and denominators independently</td>
<td>5/11 – 2/11 = 3/0</td>
</tr>
<tr>
<td>Lack of knowledge/Unfamiliarity with process</td>
<td>Adding numerators and choosing one denominator as denominator sum</td>
<td>14/15 + 2/30 = 16/30</td>
</tr>
</tbody>
</table>

Rifts spawning cause sky fragments to drop. Rifts are generated via random selection among a selection of integers. The range of the integers are determined by the observed performance of the user.

3.6 Pedagogical Component

The pedagogical component of Endless Sky uses the learner model to adjust the fraction values and whether the abstract fraction notation is shown together with the bar model or not. It decides the changes on the game component through threshold probabilities for each of the learner model’s nodes. After the learner model is updated based on the recent performance of the player, the pedagogical component issues an inference request to the learner modeling component to retrieve the updated probabilities of the learner model. The nodes’ rates (which represent the mastery of the LOs and other arithmetic skills) are used to decide how the puzzles are generated, making them easier or harder.

The pedagogical component also uses the learner model’s knowledge of the player’s misconceptions by generating “traps” that would “tempt” the player to commit an error that would expose a misconception shown in Table 4. For example, in Figure 4, the player must first solve a rift chosen by the game, which is the 2/3 rift. A set of sky fragments is generated to include some (notice the 1/1 and 1/2 sky fragments) which, when manipulated (using the LO mechanics) in a certain way (e.g. merging the 1/1 and 1/2 sky fragments without scaling them), would indicate the presence of a misconception. These traps are generated until the misconception rates are approximately 0%
Correctly solving puzzles laid with “traps” will gradually decrease the rates. Traps will not always be generated to prevent the player from becoming conscious of them.

Turning this component off causes the GILE to base its fraction value control and abstract fraction display on arbitrary game score milestones and alternating between trying a trap or not.

4. Preliminary Results

To evaluate the effectiveness of the GILE, especially the pedagogical component’s effectiveness in implementing its traps and using the learner model to control the puzzles, we planned to conduct a pretest-posttest quasi-experiment in an on-campus, in-person environment similar to what was done in (Sison, et al., 2018). However, the COVID-19 pandemic compelled us to conduct the pretest-posttests online. While waiting for a local school to conduct classes online, social media was used to gather Grade 5 or 6 participants. Participants were alternatingly assigned the active pedagogical component version and inactive version as purposively grouping participants who joined at different times was impossible. P1, P2, and P3 were the only ones that properly went through with the pretest, playtesting, and posttest.

When a local school (LHCS) began conducting a dry run of its online classes, we decided in coordination with a teacher to have the game playtested by its Grade 6 class of 19 students. Of the 19, 13 went with the pretest. The 13 were then grouped based on their pretest scores: 7 were assigned to playtest the game with the pedagogical component active (experimental group) and 6 were assigned to playtest the game with the pedagogical component turned off (control group). However, due to technical difficulties and compliance issues despite supervision of the teacher, only 4 of the 13 participants (1 from the experimental group, 3 from the control group) went through with the procedures properly.

In total, therefore, 7 participants were able to playtest the GILE and take the pretest and posttest. The results are shown in Table 6. P1, P2, and P3’s short quizzes had 8 items while P4, P5, P6, and P7 had 10 items. The difference in quiz items for the latter was at the consultation of the latter’s supervising online math teacher after she reviewed the short quizzes. For the playtesting of the GILE, participants were instructed to try beating a high score of 10,000 to get them working towards a goal when playing.

Table 6. Preliminary Pretest-Posttest Results

<table>
<thead>
<tr>
<th>Playtester</th>
<th>Pedagogical Component</th>
<th>Pretest Score</th>
<th>Posttest Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Active</td>
<td>50%</td>
<td>100%</td>
</tr>
<tr>
<td>P3</td>
<td>Active</td>
<td>37.5%</td>
<td>37.5%</td>
</tr>
<tr>
<td>P4</td>
<td>Active</td>
<td>50%</td>
<td>40%</td>
</tr>
<tr>
<td>P5</td>
<td>Inactive</td>
<td>60%</td>
<td>80%</td>
</tr>
<tr>
<td>P6</td>
<td>Inactive</td>
<td>100%</td>
<td>80%</td>
</tr>
<tr>
<td>P7</td>
<td>Inactive</td>
<td>70%</td>
<td>70%</td>
</tr>
</tbody>
</table>

Figure 5. Testing a Misconception
Of those interviewed (P1, P4, P5, P6, P7), 4 found difficulty with the game. P4 had difficulty with the Scale mechanic. P6 took 7 minutes of his 17-19 minutes of play time to familiarize with the mechanics. P1 and P5 did not specify any mechanic but found the game “confusing at first” and “a bit hard”, respectively. P7 found the game easy, scoring 57,300 points with 1 to 2 hours of play time.

P1’s pre- and posttest scores suggest that he benefited from the GILE. An online chat with his older sister told us that he played the game for 80 minutes, and, being an avid video gamer, was motivated by the desire to beat the high score, which was 10,000, by scoring 12,000. Pretest data reveal that his misconception was “Numerator and denominator are 2 independent natural numbers” (recall Table 4). This was eradicated when, while playing the game:

“he figured… that you can’t merge or cleave blocks that aren’t of the same total number right? that’s when he got it na dapat the denominators are the same before adding or subtracting”

P1 also enjoyed the game:

“he beat the 10000 points. he got 12000 at 3:20pm, he started playing at 2pm. He said he likes the app so much hahaha”

P1 also used the in-game term, “scaling”, in describing his fraction arithmetic procedure:

“I imagined scaling the 4/5 then arriving at 8/10 then subtracted since it has the same denominator so the answer is 0.” – when solving 8/10 - 4/5.

P3, P4, and P5 did not revise the numerator after changing the denominator, but only P5’s posttest showed the removal of this erroneous pattern. P5 played the game for only 20-25 minutes with a high score of 4,000, giving up due to difficulty (note that she was playing the GILE with pedagogical component turned off). While she could not recall what she did differently in the posttest, she confirmed through interview that the GILE helped her understand fraction behavior during arithmetic and the Scale mechanic was no different to how she normally deals with dissimilar fractions. Her procedure:

1. multiply the denominator
2. do the cross multiply

example:

\[ \frac{1}{3} + \frac{1}{6} \]

…

for the denominator times 6 times 3 = 18 so the common denominator is 18, 1 times 6 is 6 [and] 1 times 3 is 3, 6/18 + 3/18 = 9/18 or 1/2”

Her “cross multiply” step results in a number multiplying both numerator and denominator. This is close to how the Scale mechanic manipulates a multiplier which is used to multiply the sky fragment’s base fraction value, i.e. numerator and denominator multiplied by the same number.

P4 had difficulty with the Scale mechanic. With only 10 minutes of play time, he would not have experienced the effects of the pedagogical component. Interview confirms that he does not revise the numerator after changing the denominators and does not see any error in that.

P2, P6, and P7 show no misconceptions in their short quiz results. Any mistakes found were not recurring and other items like the ones they got wrong were answered correctly. We could not get in touch with P2. P6 and P7 showed proper standard procedures for solving dissimilar fraction arithmetic.

These findings show the potential of the GILE to improve learner performance, particularly among gamers or those who can persevere past the game’s difficulty, and those whose procedures are close to the LO mechanics.

5. Final Remarks

Fraction addition and subtraction entail arithmetic procedures that can be difficult for elementary students. In this paper, we have described a game-based intelligent learning environment (GILE), the learning outcome mechanics and other game elements of which were designed based on the literature on fraction instruction and fraction misconceptions. Preliminary results seem to point toward the ability of the GILE to improve performance and remediate misconceptions for learners who can thoroughly play the GILE and those whose procedures are close to the LO mechanics.
Our experience also reveals the difficulty of conducting quasi-experimental evaluations that are fully online. We plan to do more qualitative data collection and analysis in the near future.

Acknowledgement

The researchers would like to thank DOST-SEI and ERDT for funding this research, the playtesters who volunteered to playtest the game, artists who contributed to the game’s development, and friends who assisted in the pandemic-related difficulties that the researchers encountered.

References


Enhancing Consumer User Experience, Education and Brand Awareness through Musical Advergames

Forouzan FARZINNEJAD\textsuperscript{a}, Hadi KHEZRIAN, Mohsen KASIRI, Nilufar BAGHAEI\textsuperscript{b}\textsuperscript{*}
\textsuperscript{a}Department of Computer Engineering, Sharif University of Technology, Tehran, Iran
\textsuperscript{b}School of Natural and Computational Sciences, Massey University, Auckland, New Zealand
farzinnejad@ee.sharif.edu; Hadikhezrian@yahoo.com; mohsenkasiri181373@gmail.com; N.Baghaei@massey.ac.nz

Abstract: The effectiveness of traditional communication techniques has declined in recent years, and marketing specialists are looking at more creative ways to engage consumers. A lot of attention has recently been paid to advergames, which are seen as an attractive new marketing tool to increase consumer engagement and education of brand awareness. The term advergame refers to games that combine brand advertising with gaming to promote business products. In this paper, we study the impact of music on consumer uptake of advergames (n=900). Our results show that music in advergames is an important feature because of its capability to attract audience, enhance their user experience and increase brand awareness.

Keywords: Advergame, music, consumer engagement, user experience, education, brand awareness, megabox.

1. Introduction

Advertising is a marketing communication that employs an openly sponsored, nonpersonal message to promote or sell a product, service, or idea. Sponsors of advertising are typically businesses. Advertisements deliver the message in creative ways to reach as many audiences as possible [Ho & Lam, 2019]. \textit{"We breathe nitrogen, oxygen, and advertising every day."} says French advertising critic Robert Grange. Advertising as a cultural phenomenon has a significant impact on all aspects of our social life [Qin, 2019]. In today's advertising industry, one of the ways to increase interest in advertising is the field of entertainment programming. Digital Games as an attractive mass media have become an international marketing communication tool for advertisers [Bellman et al, 2014].

Advergaming combines a promotional message, company logo or other brand information with online games or video games. This has created an appealing and engaging platform that is becoming increasingly popular with the widespread use of social media. Advergames are regarded as innovative tools for building a customer relationship, which are both suitable for existing products and business startups [Waiguny et al, 2011]. Research findings indicate that advergames are a promising form of advertising, particularly for young people [Bellman et al, 2014]. However, there are a significant number of gamers of all ages. Also, the fact that 44% of all gamers are women shows that the dominance of males in the field of digital gaming is fading [Lee & Cho 2017]. A strategy followed by the majority of media planners these days is the consumer choice of media that enhances brand recall through entertainment. Therefore, advertising managers should also incorporate games into the design of advergames to ensure that their implementation positively influences consumers' memory [Vashisht, & Sreejesh 2015]. The Advergaming technique helps increase brand awareness and remembrance, which persuades consumers to make a purchase.

1.1 Immersion and Persuasive
Persuasive technology is defined as technology that is designed to change attitudes or behaviors of the users through persuasion and social influence [Fogg, 2002]. Such technologies are used in sales and marketing, education, politics, military training, public health, lifestyle changes, and management, and may potentially be used in any area of human-human or human-computer interaction. Most self-identified persuasive technology research focuses on interactive, computational technologies, including desktop computers, web services, mobile devices [Reddy et al, 2017], video games [Oinas-Kukkonen & Harjumaa, 2008] and more recently, virtual reality [Samarasinghe, Baghaei & Stemmet, 2020]. In spite of variations in design and appearance, an important common element in many successful computer games is the ability to attract and engage players. This is called the experience of immersion, a word widely used by gamers and critics [Jennett et al, 2008] – meaning one's entire mind is concentrated on the game as part of the user experience, aka deep mental involvement.

1.2 The Role of Music and Sound in Video Games & Advertising

Sound is an essential element in video games [Cunningham et al, 2006]. Central to UX in digital gaming is deliberate treatment of sound and music to affect the player’s feedback and rewards. In video games, audio UX is shaped by sound and audio cues. The role of audio is to facilitate interaction with the virtual gaming world and trigger interactive feedback [Nacke et al, 2010]. Because sounds relate to usability, they provide information that directly responds to or demands the player's performance. In other cases, sounds have a more informative function in which they direct the player's orientation or help the player identify different situations and situations [Jørgensen, 2008]. Background music, which is a collective term used to refer to music and sound effects, has long been considered as an inherent element of modern video games [Zhang & Fu 2015] and has proven to be an essential component of any popular video game [Areni & Kim, 1993, Politis et al, 2016]. Given that game production has become more complicated than in the past and carries high costs (up to millions of dollars), sound engineers and composers have assumed great importance.

Advertising makes use of music as a driving component in commercials to enrich the original message. The concept of environmental processing indicates that environmental cues like music can result in a good attitude towards advertising and towards the brand [Morris & Boone, 1998]. An environmental cue such as music has the most impact on brand attitudes in an advertising environment. The significant association between music and emotions has indicated that exciting music enhances emotional arousal in individuals manifested through skin reactions and increased heart pulsation, which are both physiological symptoms for emotional response. Music influences brand cognition and purchase intention in shaping and facilitates purchase intention [Morris & Boone, 1998]. Research has confirmed the profound influence of music on arousal, mood, and emotion. In fact, our brain responds to more melodies than language. Music evokes emotions, and emotions make decisions such as buying or not buying a product. So, it can be used strategically to increase sales. In general, music appears to increase cognitive load, which is an effective way to get people thinking [N.c, 2015]. Research conducted by the Nielsen Institute has confirmed that music helps to connect brands with their customers on an emotional level [Nielsen, 2019].

Experts believe that genre influences consumption habits. In a study [Areni & Kim, 1993], a wine shop began playing classical music instead of top billboard songs. As a result, it sold more expensive bottles of wine. Studying music in restaurants has also confirmed the correlation between classical music and rising costs. Consumers, therefore, are likely to have heard classical music associated with complexity and expensive tastes and have inadvertently altered their behavior to suit the environment.

In advertising, memorability is important - more effective advertising stays in memory. Although consumers are expected to remember the message of an advertisement, this will directly lead to shopping only if accompanied by emotional involvement. In fact, the best advertisements are those which are both informative and emotionally powerful. When an advertisement uses music, the audience strives to identify and remember the original message, which in turn improves business sentiment and customers buy from their favorite brand [N.c, 2015].

Our project examines the effectiveness of providing music on consumers’ uptake of advergames and enhancing their user experience. In this paper, we outline the design, implementation, findings and plan for the future work.
2. Implementation & Evaluation Study

We conducted a study at the International Electronic, Computer & E-Commerce Exhibition, Tehran, Iran. It is the greatest commercial event in Iran’s Market of Electronics and Computer Products and services to advertise the first mobile brand, the world's first and largest mobile and high-speed network operator in the country. The target audience of the games were people of different age and gender.

The Megabox [Boks, 2020] machine is an advergame display device that has an industrial monitor that is vertically positioned. On the screen of the monitor is a touch module so that the user can interact with it. The coordinates (x, y) of a point on the touch module are sent depending on what point the user touches. The device also includes a speaker for audio playback. All devices are connected to the Internet so that they can text the bonus code to the user or print it on paper using a printer embedded in the device. To identify and prevent participants from playing a game more than once, they were asked to enter their phone numbers into megabox.

![Figure 1. Megabox mechanism](image)

This game we chose for this study was adopted from Piano tiles [Wikipedia, 2020]. The participants first selected one of the activities and services related to the brand's in-app icons, including charging purchases, Internet pack purchases, etc. By selecting the charging icon, the service provided features such as inserting a charging card, charging a purchase, charging purchase history and custom charging. Instead of black tiles in the original piano tiles game, the services were presented in circle tiles. When each services circle tile is tapped by the user, it will make a sound of cooperate music we intend to promote instead of piano tile music. If the player taps on a white tile, the player will lose the game and be signaled by an off-tune note. Similar to Piano Tile, the game had three levels of speed, and after touching a certain number of circles, the game speed increased. It should be noted that touching each of these circles increased the player’s score by one. As the participant’s score increased, more valuable prizes remained and fewer valuable prizes were eliminated. The purpose of this scenario was to convey the intended advertising message (brand features of the app) by advergame.

A chance mechanism was also designed to reward participants. Sixteen bonus boxes appeared with more valuable prizes depending on the player's score in the game. The participants selected one of the remaining bonus boxes, and the prize would appear. The participants were able to receive their prize by presenting the brand name printed by Megabox to the brand booth. For the study, we used two Megabox machines, with two different versions of the game (Piano advergame with and without background music). The numbers of females and males and the participants' age are as follow:

<table>
<thead>
<tr>
<th>Table 1</th>
<th>The number of females and males</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Females</td>
</tr>
<tr>
<td>Piano tiles with background music</td>
<td>48%</td>
</tr>
<tr>
<td>Piano tiles without background music</td>
<td>32%</td>
</tr>
</tbody>
</table>
During the gameplay, the services, features, new features of that company's application, product or brand were passed on to the target audience, depending on the advertising part of an advergame, and nature and need of the brand. At the end of a session, according to the game's aspect in the advergame and the need for a reward, the player was rewarded by the brand, proportion to their score.

3. Results

The results from a Chi-square test show that considering gender and age group, there is no significant difference between the two groups (p>0.05). According to product evaluation by individuals, satisfaction with advertising game, possibility of introducing to friends, and satisfaction with experience, there is a significant difference between two groups (advertising game with music and advertising game without music) (p<0.05). Descriptive results show that those who chose advertising game with music at 95% of confidence, have had a more positive evaluation of the product than those who chose advertising game without music. Also, satisfaction with advertising game in the first group (advertising game with music) at 95% of confidence, is higher than the second group (advertising game without music). Possibility of introducing to friends in the first group (advertising game with music) at 95% confidence is higher, comparing to the second group (advertising games without music). Also, satisfaction with experience in the first group (advertising game with music) at 93% of confidence is higher than the second group (advertising games without music).

Table 1: Descriptive (Frequency and Percentage) and inferential Chi-Square Test results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Group</th>
<th>Advertising Game with music</th>
<th>Advertising Game without music</th>
<th>Chi-square Test results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Very Satisfied</td>
<td>53</td>
<td>8</td>
<td>X²=49.274, P=0.000</td>
</tr>
<tr>
<td></td>
<td>Very Dissatisfied</td>
<td>3</td>
<td>9</td>
<td>1-0.05</td>
</tr>
<tr>
<td></td>
<td>Satisfied</td>
<td>39</td>
<td>13</td>
<td>3-0.05</td>
</tr>
<tr>
<td>Mass</td>
<td>Dissatisfied</td>
<td>9</td>
<td>27</td>
<td>15-0.05</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>20</td>
<td>23</td>
<td>16-0.05</td>
</tr>
<tr>
<td>Total</td>
<td>124</td>
<td>100/0</td>
<td>73</td>
<td>100/0</td>
</tr>
<tr>
<td>Product evaluation by individuals</td>
<td>Very Satisfied</td>
<td>106</td>
<td>27</td>
<td>X²=49.274, P=0.000</td>
</tr>
<tr>
<td></td>
<td>Very Dissatisfied</td>
<td>18</td>
<td>46</td>
<td>1-0.05</td>
</tr>
<tr>
<td></td>
<td>Satisfied</td>
<td>124</td>
<td>73</td>
<td>100/0</td>
</tr>
<tr>
<td>Satisfaction with advertising games</td>
<td>Yes</td>
<td>2</td>
<td>11</td>
<td>X²=65.909, P=0.000</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>106</td>
<td>27</td>
<td>0-0.05</td>
</tr>
<tr>
<td>Total</td>
<td>124</td>
<td>100/0</td>
<td>73</td>
<td>100/0</td>
</tr>
</tbody>
</table>

Figure 2. Left to right: activities and services related to the brand, the service provided features, the game, chance mechanism, the prize
After 10 days, the participants were asked again if they remembered the game. The difference between the two groups is statistically significant ($p<0.05$). The result shows that the proportion of *not remembering* variable in the first group (advertising game with music) is just 4 percent, while the value of this variable in the second group (advertising games without music) is 26%. 6.5% of individuals in the first group remembered the company but not product or advertisement, while this value was 19.2% for the second group. 16.1% of the first group and 30.1% of the second group remembered the company and product but not the advertising game. Finally, 73.4% of the first group remembered the advertising game, while this value was 24.7% for the second group.

Our results also showed that the audience were more likely to play the Piano tiles with background music than without.

---

**Table 1.** Number of participants in Piano tiles with and without background music over three days

<table>
<thead>
<tr>
<th>Time</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>10/2/2020</td>
<td>32</td>
</tr>
<tr>
<td>11/2/2020</td>
<td>55</td>
</tr>
<tr>
<td>12/2/2020</td>
<td>37</td>
</tr>
<tr>
<td>Total</td>
<td>124</td>
</tr>
</tbody>
</table>

**Table 2.** Number of participants in Piano tiles without background music over three days

<table>
<thead>
<tr>
<th>Time</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>10/2/2020</td>
<td>17</td>
</tr>
<tr>
<td>11/2/2020</td>
<td>35</td>
</tr>
<tr>
<td>12/2/2020</td>
<td>21</td>
</tr>
</tbody>
</table>
| Total      | 73           

*Figure 3a & 3b.* Number of participants in Piano tiles with and without background music over three days
4. Discussions & Future Work

We conducted a study to examine the uptake of a game with or without background music, in terms of the number of participants in an advergame. Our findings show that music in advergame is effective in attracting the audience and enhancing their user experience. Going forward, we will add personalisation to the games to see if that makes them more persuasive in engaging customers and enhancing their education of the brand awareness.

Acknowledgements

We would like to thank Boks corp. for sharing their data and Dr Atefeh Ahmadi for her contribution to this project.

References

Engaging the families with young children in museum visits with a mixed-reality game: A case study

Tien-Yu Hsu, Hsin-Yi Liang*, Jun-Ming Chen

Department of Operation, Visitor Service, Collection and Information Management, National Museum of Natural Science, Taichung, Taiwan

*cidy@mail.nmns.edu.tw

Abstract: Museums provide open-ended learning environments with fruitful learning resources for children. Most museums view children as members of the family groups because children usually visit museums together with their caregivers. Therefore, there is a need to facilitate the caregiver-child interactions and promote their participatory experiences for enhancing their re-visiting motivations. To this end, this study developed a mixed-reality game to promote social interactions and participation for family groups with young children. Preliminary research was conducted to understand the learning effectiveness of the mixed-reality game. Two family groups with preschool children were invited to experience the game. Qualitative interviews were applied to understand the learning performance of young children and the feedback of both caregivers and children. Besides, their learning behaviors were observed and conversations analyzed with a lag sequential analysis to investigate the differences in social interactions. The results showed that the mixed-reality game can enhance children’s learning in museums. Besides, questioning can facilitate more social interactions for children. The results can be used for future improvements to better engage the family groups with young children in museum visits.

Keywords: Museum learning, family group, preschooler, social interaction, game-based learning

1. Introduction

Museums provide open-ended learning environments with fruitful learning resources for children. The Contextual Model of Learning (CML) indicates each child’s learning preferences to decide their learning behaviors during museum visits; their social interactions and interactions with museums’ physical environments shape their museum experiences and motivations for return visits (Falk & Dierking, 2000). Similarly, the Sociocultural Theory of Development proposed by Vygotsky (1978) explains that children learn through their interactions and conversations with others. The interactions between children and others can deepen their impressions and motivations for museum learning. The highly-interactive experiences will become good memories for the children, so that when they become adults, they will be encouraged to bring their children for return visits (Black, 2005; Dockett, Main, & Kelly, 2011).

Many studies highlight the key role of interactive experiences to support children’s learning (Donohue, 2014; Henderson & Atencio, 2007). However, previous research also indicates consideration of the needs of both adult caregivers and children to effectively support children’s learning in museums because most children do not visit museums on their own (Falk & Dierking, 2018). Determining how to facilitate meaningful communication and interactions between caregivers and children has become one of the important issues for worldwide museums. However, many visitors in family groups do not have sufficient inquiry skills and interest in exhibitions; in turn, they seldom observe the exhibits deeply or have meaningful conversations regarding the museum exhibitions (Gutwill & Allen, 2017). Besides, many museum services are still designed mainly for children; it is difficult to satisfy the different needs of adult caregivers and effectively promote adult caregiver’s engagements for meaningful caregiver-child interactions (Degotardi, Johnston, Little, Colliver, & Hadley, 2019).
To promote social interactions between caregivers and children, researchers have indicated that questioning is an effective way to facilitate meaningful inquiries and social interactions between caregivers and children (Haden et al., 2014). Asking questions can facilitate the attention paid to exhibits for family groups, and provide a scaffold for them to observe the exhibits deeply and discover more related knowledge regarding the museum exhibitions (Gutwill & Allen, 2017).

In addition to questioning, previous research indicates that active participation in learning with social interactions can effectively enhance children’s understanding and imprint what they learned during their engagements (Benjamin et al., 2010). To promote the active participation in museum learning for family groups, various interactive technologies have recently been applied in museum learning to build a mixed-reality (MR) environment with immersing experiences, such as interactive exhibitions, augmented reality, and virtual reality (Ch’ng, Cai, Leow, & Zhang, 2019; Chiang, Yang, & Hwang, 2014). The MR environment combines virtual worlds in real-world contexts, and the visualizations of concepts support learner’s constructions of abstract concepts in an MR environment (Frank & Kapila, 2017). Previous research also shows that MR can significantly improve learning motivation, active participation and enjoyment (Raptis, Fidas, & Avouris, 2018). Besides, it offers the potential to facilitate more novel experiences and rich interactions between adult caregivers and children because innovative technology can change the traditional social roles of knowledge providers for adults and receivers for children, bringing new surprises for both adults and children during their unknown explorations (Degotardi et al., 2019).

In this vein, this study developed a GBL service with MR technologies and puzzles to engage the family groups with preschool children actively interacting with museum exhibits, and facilitate rich social interactions between adult caregivers and children. Preliminary research was conducted to understand the learning effectiveness of the MR game. The results will be used for future improvements in museum services or GBL applications and to facilitate rich social interactions for family groups to motivate return visits. In brief, this study intends to answer the research questions:

- Can MR game support young children’s learning in museums?
- How did the family groups’ learning with the MR game in museums?
- How can MR game facilitate meaningful social interactions for family groups?

2. System design and learning scenarios

To engage the family group with young children in their museum visits and facilitate meaningful social interactions, a mobile GBL system was developed and implemented in the National Museum of Natural Science (NMNS). Location-based augmented reality technologies were used to identify the locations of museum exhibits and the family groups to provide MR interactions during the gaming process. An MR game named “Fighting for color” was designed for the family groups with preschool children to learn about the animals on earth, as well as the names and functions of animals’ colors.

There are two puzzles in this game. In the beginning of the game, a black-and-white tiger cries for his lost color. The children are asked to find the missing color for the tiger; they were able to find the exhibit of the tiger through the MR clue. Once the children find the specific exhibit, the supplementary learning materials include audio clips and images which can help the children to learn more about the animals and their colors, such as the names and roles of the animals’ colors in the environments. After the children finish the puzzle, another puzzle is given to ask children to find “the animal with black and white stripes (zebra).” The MR clue is provided to guide the children to visit where the zebra exhibit is found, and an introduction on the zebra and its colors is provided after they find the exhibit area. Figure 1 displays the system flow of this MR game.
As shown in Figure 2, the children were asked to play the game with their caregivers; they were able to discuss and find the exhibits together during the gaming process. The puzzles encourage the family group to explore the museum and observe the museum exhibits to learn about the animals and their colors, such as orange, red, and black. The supplementary learning materials support the adult caregivers in interacting with their children for enjoying more conversations. For example, the learning summary explains the function of animals’ colors, and the adults can support the children extended learning or discover other related exhibits together. The total game duration is about 15 minutes.

3. Experiment design

To answer the aforementioned research questions, a case study was conducted in the NMNS. Two family groups were invited to experience this MR game. One group was a female with a three-year-old boy, and another group was a female with a four-year-old boy. Both groups often visit museums in their leisure time, and they have experiences of learning with mobile devices. The children can follow the instructions and obtain a basic understanding of animals.

To understand how the family groups interact with MR, the learning behaviors and conversations of family groups were recorded by a video recorder. Regarding the social interactions between caregivers and children, each sentence of their dialogues was analyzed by the dialogue code scheme based on the parent-child interactions as observed by Degotardi et al. (2019). Table 1 shows the dialogue code scheme.
Table 1. The dialogue code scheme

<table>
<thead>
<tr>
<th>Role</th>
<th>Code</th>
<th>Definition</th>
<th>Descriptions</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caregiver</td>
<td>PA</td>
<td>Arousal attention</td>
<td>The caregiver uses verbal sounds or gestures to raise the children’s attention.</td>
<td>Look! A tiger!</td>
</tr>
<tr>
<td></td>
<td>PI</td>
<td>Information</td>
<td>The caregiver reads the descriptions of the exhibits or MR game.</td>
<td>The tiger said his color was stolen by an evil magician.</td>
</tr>
<tr>
<td></td>
<td>PT</td>
<td>Instruction</td>
<td>The caregiver instructs the child to follow the instructions.</td>
<td>Click this button.</td>
</tr>
<tr>
<td></td>
<td>PQ</td>
<td>Question</td>
<td>The caregiver asks questions to facilitate children’s responses.</td>
<td>Where is the tiger?</td>
</tr>
<tr>
<td></td>
<td>PW</td>
<td>Tutorial</td>
<td>The caregiver explains or teaches a new concept for children.</td>
<td>This is a tiger.</td>
</tr>
<tr>
<td></td>
<td>PM</td>
<td>Recall</td>
<td>The caregiver links the experiences to previous experiences.</td>
<td>Can you remember when we saw a zebra last year in a zoo?</td>
</tr>
<tr>
<td>Child</td>
<td>CA</td>
<td>Arousal attention</td>
<td>The child makes sounds or gestures to raise the caregiver’s attention.</td>
<td>Mommy! It’s a tiger.</td>
</tr>
<tr>
<td></td>
<td>CQ</td>
<td>Question</td>
<td>The child asks questions.</td>
<td>Why did the zebra steal the tiger’s colors?</td>
</tr>
<tr>
<td></td>
<td>CE</td>
<td>Explanation</td>
<td>The children express their ideas or answers to a question.</td>
<td>Stealing is bad.</td>
</tr>
<tr>
<td></td>
<td>CM</td>
<td>Recall</td>
<td>The children link the current experiences to previous experiences.</td>
<td>Mommy, we saw a tiger before.</td>
</tr>
<tr>
<td></td>
<td>CF</td>
<td>Emotion</td>
<td>The children express their feelings and emotions.</td>
<td>I feel so sad.</td>
</tr>
</tbody>
</table>

To understand the learning performance of preschool children and the feedback of family groups, deep interviews were conducted before and after the experiment. The family groups were asked to play the MR game and finish the two puzzles together. After they finished the game, an in-depth interview was held to evaluate the learning performance of children and the family group’s perceptions regarding the MR game. The total duration was about 30-40 minutes.

4. Results and discussions

Table 2 displays the dialogue codes from both family groups. The results indicated that the caregivers played the main role of facilitating social interactions in the family groups. In the first group, the caregiver was the main operator of the mobile device, and the caregiver and child watched the tablet screen together. In the second group, the main operator was the child, and the caregiver supervised his manipulation; they watched the tablet screen together. The results showed that the adult caregivers usually instructed the children to follow the instructions during the learning process (PT). Besides, the caregivers also used questioning to facilitate children’s expressions (PQ). Conversely, the children seldom asked questions during the gaming process, and usually explained their ideas and responded to the caregivers (CE). The results also showed that the caregiver and the child of the first group had more social interactions in comparison with the second group. The reason may be because the child of the first group did not manipulate the tablet, and the caregiver used more social interactions to share the attention with the child and facilitate the child’s engagements (Degotardi et al., 2019). Besides, the child in the first group also used more social interaction to keep engaging in the MR game.
### Table 2. The frequency and percentage of dialogue codes of family groups

<table>
<thead>
<tr>
<th>Group</th>
<th>PA</th>
<th>PI</th>
<th>PT</th>
<th>PQ</th>
<th>PW</th>
<th>PM</th>
<th>CA</th>
<th>CQ</th>
<th>CE</th>
<th>CM</th>
<th>CF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group1</td>
<td>1</td>
<td>7</td>
<td>14</td>
<td>11</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>16</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1.92%</td>
<td>13.46%</td>
<td>26.92%</td>
<td>21.15%</td>
<td>3.85%</td>
<td>0%</td>
<td>0%</td>
<td>1.92%</td>
<td>30.77%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Group2</td>
<td>0</td>
<td>7</td>
<td>8</td>
<td>13</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>17.07%</td>
<td>19.51%</td>
<td>31.70%</td>
<td>4.88%</td>
<td>4.88%</td>
<td>0%</td>
<td>0%</td>
<td>19.51%</td>
<td>2.43%</td>
<td>0%</td>
</tr>
</tbody>
</table>

To further investigate the social interactions between the adult caregivers and children, log sequential analysis was applied to analyze the social interaction patterns of both family groups. A Z value greater than 1.96 reached the level of significance ($p<0.05$). As shown in Figure 3, the results indicate the different social interaction patterns of both family groups. In the first group, the caregiver usually asked questions, and the child responded and expressed his ideas (PQ $\rightarrow$ CE). In the second group, the caregiver linked the experiences to previous experiences, and provided related knowledge for the child (PM $\rightarrow$ PW). Besides, the recall of previous experiences encouraged the child to express his previous experiences (PM $\rightarrow$ CM). After the child expressed his ideas, the caregiver asked related questions to facilitate more interactions (CE $\rightarrow$ PQ).

![Figure 3. The social interaction patterns of family groups](image)

Regarding the learning performances, the results of the interview showed that the children in both groups could identify the animals and colors after the learning activity. Hence, the MR game can support young children’s learning in museums. Regarding the perceptions of the MR game, both caregivers and children think that the MR game is a playful way to engage them in museum visits. The caregivers appreciated the MR game can facilitate the children’s explorations and understanding of museum exhibitions. However, there are some issues to effectively support children learning with MR games. Firstly, the manipulations and learning contents of MR games directly influenced the family group’s learning experiences. Because the preschool children cannot read and understand the text, playing the MR game still rely on the caregiver’s operations and instruments. To improve the active participation of young children, voice-control manipulation design and graphic interfaces can be used to assist the young children interacting with MR games. Another issue is that young children’s height is usually lower than the museum exhibits, and it is hard for young children to scan the exhibits independently. Hence, there is a need to build a friendlier learning environment for young children’s learning in museums.

## 5. Conclusions

This study developed an MR game to engage family groups with young children to learn in museums. The results showed that the MR game can support the young children learning in museums and facilitate their social interactions and engagement. The caregivers usually play the role of main facilitator of social interactions by asking questions and providing instructions; there are fewer children’s
interactions compared to the caregivers. The results indicated a need to encourage more children’s interactions and enhance their reflections on learning. To improve young children’s learning and increase their social interactions, the future design of MR games can add more questions related to the children’s previous experiences and encourage more reflections on learning.

This study carried out an empirical exploration of social interactions between caregivers and young children in museum learning. The results can be used to design effective interactive learning services for family groups and facilitate rich social interactions. More family groups can be included in future studies. Future studies can be conducted to analyze the learning behaviors regarding the interactions with museum exhibits and MR games, and the children’s knowledge constructions in connection with the social interaction patterns. In addition, the effect of different MR game design on children’s learning and the social interactions between the caregivers and the children can be investigated to provide more effective museum services for family groups learning in museums.

Acknowledgments
This work was supported by the National Science Council of Taiwan, R.O.C., under contract numbers MOST 108-2511-H-178-001.

References

Applying Component-based Strategy to Design an Educational Board Game for Children to Learn Chinese Characters

Hui-Ling HSU* & Zhi-Hong CHEN*
*Graduate Institute of Information and Computer Education, National Taiwan Normal University, Taiwan
*huilin@jcps.tp.edu.tw

Abstract: This paper proposes an educational board game, “Chinese-character-monsters” (CCM), to facilitate primary students in Chinese character learning. Each monster game card represents a specific component of Chinese characters, which can be further combined as various Chinese characters. In the board game, students are offered a number of opportunities to apply this component-based strategy, they thus can be more aware of how Chinese characters consist of different components. A preliminary evaluation was also conducted to investigate its impacts, including satisfaction, self-efficacy, and flow experience. The results showed that the CCM educational board game had a positive effect on students. These findings will guide our design to further modify and tune the CCM board game.

Keywords: Chinese character, radical, component, educational board game

1. Introduction

Chinese is considered as one of the hard-to-learn language systems. This is because Chinese characters are non-alphabetic in form, and students need to learn various characters. In addition, since Chinese characters consists of whole characters, components and strokes, such structure complexity often confuses students, especially when they learn the characters with the same pronunciation but different meanings. In spite of understanding the meaning of the words, students are not always write Chinese characters correctly. The familiarity of the radicals and the components are important in the process of Chinese character learning. However, the lower graders may treat characters as unanalyzed wholes so that they cannot systematically make use of the components of characters as skilled Chinese learners (Shu & Anderson, 1997).

Chinese morpheme usually consists of characters based on a word, a radical, or a character. Component (bù jiāns, 部件) are the minimal orthographic unit in Chinese writing. They can be as small as a stroke, or as large as a semantic radical. Multiple orthographic structure in Chinese characters often causes learners’ intrinsic cognitive load. Previous studies have shown that understanding radicals is crucial for learning Chinese characters (Chen, Hsu, Chang, Lin, Chang, Sung, 2013; Huang, 2006). A radical is the core of the components from which a character is identified. Some researchers indicated that this orthographic knowledge can be delivered explicitly by teachers who use the radical-based character teaching method (Chen et al., 2013).

The traditional form of Chinese characters, which is used in Taiwan and can have several strokes and components, were overly complex for learners. There are about 5000 high-frequency Chinese characters, with a total of 440 components, half of which are radicals. In other words, to learn Chinese characters well, you must not only know the knowledge of radicals and familiarize with the components to expand literacy and strengthen the learning effect (Huang, 2003). Learners need to gradually learn Chinese characters by recognizing and compounding radicals and components correctly.

The concept of game-based on Chinese language learning is not new. Game-based learning design is not only presenting teaching content in games, the learning objectives are also critical. However, the advantages of board games are convenient and easy to get started, it is suitable for using in teaching especially in elementary school classes that contain young children. Earlier studies mostly...
emphasized radical-based Chinese characters learning strategies. In this study, the researchers aim to develop the learning aids with an increased focus on component-based design to help learn Chinese characters. Therefore, it is necessary to further examine the strategy’s feasibility and applicability.

2. Chinese Characters Board Game Design

2.1 Chinese character learning

Learning to recognize Chinese characters is one of the most difficult learning items for learners. Past studies have pointed out that students could maintain stronger motivation for Chinese characters learning with gaming approaches. In addition, radicals become even more critical in literacy because of the semantic clues they provide for characters. For orthographic processing, both children and adults make use of radicals when reading compound characters, positional information is critical for activating radical information during character recognition (Lin, Wang & Singh, 2018). Previous research has developed a game-based model for Chinese character learning, named Character-Monster (Chen, Chi & Ciou, 2017). Compared with digital game systems, board games are more portable and versatile, easier to use in the class. Therefore, this study developed a component-based Chinese character board game for elementary school children in Chinese course.

2.2 Chinese Character-Monster (CCM) board game

“Character-Monster” is an educational board game using component-based strategy to learn Chinese characters for learners from 6 to 9 years old. Learning vocabulary by visualization is a method using images to present and introduce the formation of characters and the development of the form of characters. It also explains and analyzes the characters having similar structures, sounds or meanings. This method is based on the theory of etymology and cognitive psychology, and could help learners recognize words faster and have a deeper impression (Chou, 2009). Familiar with the rules of radical position is helpful in the process of learning Chinese characters (Chen, Chang, Chiou, Sung & Chang, 2011; Hong, Wu, Chen, Chang, and Chang, 2016; Zhan & Cheng, 2014). Accordingly, in this study, the basic level game includes 51 high-frequency components and 86 high-frequency characters to assist learners in developing radical awareness and character processing.

The gameplay of the system includes three phases, “Combining Components”, “Making Words” and “Count Attack Points”. Each of Character-Monsters cards represents a corresponding component and points. Apart from this, 6 magic cards are also designed to increase the fun and excitement in this game. Figure 1 shows an example of a monster card and a magic card design.

![Figure 1. The Hand Cards Surface Design](image-url)
Learners are randomly assigned as a group of four members. Figure 2 shows three phases of the CCM board game. The players are firstly given by five cards. In the Combining Components phase, players combine the cards with different components to form a character. Players take turns attacking. In the Making Words phase, the attacker gives the words based on the character to enhance their attacking power. After the attacker completes the two phases, the defender can decide whether to use the magic card to resist attacks. At last, both players add up the total points on the card used. If the sum is greater than the opponent's points, then the player wins. After a round, each player are given three more cards. The game is over until all the cards are used, the one who got the most points is the winner.

![Figure 2. The phases of the CCM board game](image)

3. Analysis

3.1 Participants

A total of 20 second graders from an elementary school in northern Taiwan, participated in this study, with a mean age of 7.6. Of these participants, 9 (45%) were males, 11 (55%) were females. They are all voluntary participants and also include 2 attention-deficit children. After the participants completed the game, they were asked to fill in a questionnaire to collect their experience. Participants all have experiences in playing card board games, but none have used the CCM board game before, and statement that they usually learn characters by writing repeatedly.

3.2 Questionnaire

As the participants are in young ages, the teacher had explained the questions one by one and confirmed that everyone understood the meaning of the questions clearly before answering while they were filling out the questionnaire. Do these steps sequentially until the entire questionnaire was completed.
The questionnaire contains 12 items, which are composed of three dimensions to present a level of perception (satisfaction, self-efficacy, flow-experience) while playing the CCM scaled by 5-points Likert format (1=Strongly Disagree to 5=Strongly Agree). The items is developed based on the definition of Quesenbery (2004), engaging is how pleasant, satisfying, or interesting an interface is to use. To evaluate the consistency of the variables, a reliability analysis of the questionnaire is identified using Cronbach’ α. According to George & Mallery (2003), a Cronbach’ α value above .7 indicates acceptable reliability. As shown in Table 1, this study used SPSS as an analytical tool to carry out descriptive statistical analysis. The means of the satisfaction were between 4.5 and 4.8; the means of the self-efficacy were between 4.65 and 4.75; the means of the flow-experience were between 4.55 and 5.

Table 1. Perception Analyses

<table>
<thead>
<tr>
<th>Construct</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Satisfaction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. I think the cards of this board game is well designed.</td>
<td>3</td>
<td>5</td>
<td>4.8</td>
<td>.523</td>
</tr>
<tr>
<td>2. I think the appearances of character monsters deepen my impression of radicals.</td>
<td>4</td>
<td>5</td>
<td>4.8</td>
<td>.410</td>
</tr>
<tr>
<td>3. I think it is easy to operate by combining.</td>
<td>3</td>
<td>5</td>
<td>4.5</td>
<td>.688</td>
</tr>
<tr>
<td>4. Character monster cards help me acquaint words.</td>
<td>3</td>
<td>5</td>
<td>4.7</td>
<td>.657</td>
</tr>
<tr>
<td><strong>Self-efficacy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. I think the cards of this board game is helpful when acquainting components (radicals).</td>
<td>4</td>
<td>5</td>
<td>4.75</td>
<td>.444</td>
</tr>
<tr>
<td>2. I think it will deepen my impression of character component structures through combining.</td>
<td>4</td>
<td>5</td>
<td>4.75</td>
<td>.444</td>
</tr>
<tr>
<td>3. I combine characters faster after using this board game.</td>
<td>3</td>
<td>5</td>
<td>4.75</td>
<td>.550</td>
</tr>
<tr>
<td>4. I would be able to recognize and pronounce the words that appeared on the cards correctly when I see them on my quiz.</td>
<td>3</td>
<td>5</td>
<td>4.65</td>
<td>.671</td>
</tr>
<tr>
<td><strong>Flow-experience</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. I don’t feel like taking breaks every time I play this board game.</td>
<td>4</td>
<td>5</td>
<td>4.95</td>
<td>.224</td>
</tr>
<tr>
<td>2. I think this board game is fun and I want to keep on challenge.</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>3. I hope to often use this board game to practice acquainting characters.</td>
<td>4</td>
<td>5</td>
<td>4.8</td>
<td>.410</td>
</tr>
<tr>
<td>4. I was immersed in learning Chinese characters while using the CCM board game.</td>
<td>3</td>
<td>5</td>
<td>4.55</td>
<td>.686</td>
</tr>
</tbody>
</table>

3.3 Interview

In addition to the questionnaire, the researcher also interviewed the participants and asked them several questions such as "Do you have any suggestion for the CCM board game designer?", "Which part of the CCM do you like the most?", "What do you think CCM board game can help with your learning?" and "How do you think about your performance in the CCM game?", and ask for further details based on their responses promptly to collect qualitative data about users' perception. In reporting the results of the study ‘P1’ stands for ‘Participant 1’, ‘P3’ stands for ‘Participant 3’. Some participants expressed the hope that researchers can provide characters and words lists so that they can learn by themselves before the game. The opinion was echoed by P1, P4, P6, P12, P13 and P20. Two of them even hope to take a set of CCM board game card home for practice. In summary, the participants expressed high level of engagement and positive feedback on the CCM board games, but they also reflected the need to practice CCM content at home individually. The participants’ comments were as follows:

- "Where can I buy this board game, I want to practice at home?" (P8)
• “If there is a Game Manual, we can get familiar with the characters and words on the CCM cards, and it could be more fluent when playing.” (P9)

In addition, participants self-reported positive feelings and attitudes toward the CCM board game. They said the game was easy to learn and they are satisfied with the design of the cards. They thought the game was helpful for recognizing components or radicals to learn Chinese characters. They enjoyed playing the game. Thus, the game mechanism of CCM is satisfactory to participants and can also arise their interest. Some of the interview results are as follows:

• “The monsters on the card are so beautifully painted, so I would like to play it [CCM board game] more.” (P3 and P12)
• “I am very eager to play it [CCM board game] again soon.” (P2, P9, P13, P15 & P18)
• “When playing, I always need to see if the characters can form a word. I watch it carefully and keep it in my head because I want to win.” (P1)

4. Conclusion and Further Work

Prior research (e.g. Chen et al., 2013; Rawendy, Ying, Arifin, & Rosalin, 2017; Wen, 2018) has reported that each interactive digital Chinese character learning strategy has certain features which facilitate particular educational approaches. In the case of component-based learning strategies for Chinese characters, this study explored the design of CCM character board game. The survey results showed that the participants enjoyed it and looked forward to the next challenge. This study concluded that creating interesting learning experience for children must create an attractive learning tools to motivate children first. Using learning tools like CCM board game give a different choice from traditional learning strategy like writing repeatedly. The greatest value of this research lies in its demonstration effect on the design mechanism and application value of component-based Chinese character board game. It is better that the researcher can conduct experiments in a larger range and further conduct achievement tests to evaluate learning effectiveness.

Despite positive feedback on gameplay, how to improve the impact of this game on expanding words is one of the key items for future study. For example, providing a player’s word list or supplementary teaching materials, which can not only provide a game scaffold, but also help players expand their words. In the future, the researcher will further revise and optimize the game content and design based on the experimental results.

Some limitations of the study should be noted. The participants were chosen from active volunteers, but it lacks the inference of random samples. Besides, the generalizability of the outcomes may be limited because only 20 volunteers were tested.

Acknowledgements

We would like to thank all the people who prepared and revised previous versions of this document; the principal, teacher and volunteers of our partner public school for their participation.

References


Game-Based Learning in Language Education: A Review of Empirical Studies from 2009 to 2018

Chang XU, Xian ZHANG, Xiaohan YU, Yutong LU, Chunping ZHENG*
School of Humanities, Beijing University of Posts and Telecommunications, China
*zhengchunping@bupt.edu.cn

Abstract: This study conducted a systematic review of 35 empirical research regarding game-based language learning (GBLL) published by five high-impact international journals (Language Learning & Technology, ReCALL, Computer Assisted Language Learning, System and CALICO) from 2009 to 2018. Based on content analysis, we analyzed the research trends, research foci, and strengths and challenges of the game-based language learning. A number of findings were reported based on the above three aspects. The results showed that GBLL improves the overall language proficiency and academic performance of learners. Challenges were summarized for better designing and developing future digital games for language education.

Keywords: Digital games; Language teaching; Systematic literature review; Content analysis

1. Introduction

Employing digital games for the purpose of language learning has become ever more pertinent in computer-assisted language learning research within the last decade. Digital games themselves have many characteristics that are expressly applicable to general learning contexts, and specifically those found in the domain of L2 learning (Sykes & Reinhardt, 2013). In this study, three main questions are used to guide the analysis of the literature.

a) What is the general research trend indicated by the GBLL studies?
b) What are the main research foci of the GBLL publications?
c) What are the promises and challenges of GBLL revealed by the publications?

2. Methodology

The content analysis method was adopted in this study. The collected data was analyzed according to the coding framework by Zheng et al. (2019) and then presented with frequencies along with detailed descriptions.

In this literature review, five representative refereed journals in the field of technology-enhanced language learning, Language Learning & Technology (LLT), ReCALL, Computer Assisted Language Learning(CALL), System and CALICO, were selected to search for research articles on GBLL. First, research papers were chosen for containing words like “game”, “gaming”, “play” and names of games in their titles, keywords and abstracts through reading every article in the five journals from 2009 to 2018. After searching, 42 articles related to game-based language learning were selected. As a few articles did not employ digital games and/or did not connect digital games with language learning, the 42 articles were re-examined carefully. Finally, seven unrelated articles were removed and 35 articles were identified for further analysis.
3. Results

3.1 General research trends of the empirical studies from 2009 to 2018

3.1.1 Number of empirical studies published from 2009 to 2018

Thirty-five empirical studies that employed digital games to promote students’ language learning in the five journals during 2009-2018 were identified for review. In the past ten years, ReCALL and CALICO published the largest number of studies (12 and 9 studies respectively), accounting for more than half of the total. Studies focusing on the application of digital games in language learning has gradually increased since 2011, with certain ups and falls in attention. The year of 2012 witnessed the largest number of publications, in which, six studies were published in five journals. In general, the total number of articles published in this field is still quite small and the empirical studies were still in the preliminary stage.

3.1.2 Research sites of the empirical studies from 2009 to 2018

As we can see from Figure 1, more than one-third studies were conducted in America. In general, most of the empirical studies on digital game-based language learning were carried out in major developed countries, while relevant studies in other countries were insufficient.

3.1.3 Target languages of the empirical studies from 2009 to 2018

Figure 2 depicted the target language of the empirical research during this decade. During the past decade, English was the most important target language, which reflects that English as the lingua franca of the world is widely used in academic, economic and technical fields. It also shows using emerging technologies to learn English well is still a vital topic in the field of language teaching, which has great value and potential to be further discussed.

Hong et al. (2017) investigated the effectiveness of using Chinese radical learning game (CRLG) to practice recognizing Chinese radicals. With China increasingly playing its role on the global stage, Chinese is becoming more and more popular, and studies on Chinese learning through digital games is expected to increase in the next few years.

3.1.4 Research settings of the empirical studies from 2009 to 2018
As indicated in Figure 3, more than half of the articles we reviewed were conducted in high education (undergraduate and postgraduate) settings, indicating the popularity of applying the digital games to the language education among adult students. It is easy to understand since adult learners may be more competent with new technological innovations such as virtual reality employed in some digital games. An interesting research trend is that researchers have started to realize the potential effectiveness of applying digital games in language learning.

As Pitura and Pacut (2018) mentioned, being able to communicate in foreign languages, along with the remaining key competences for lifelong learning, are of special significance for upper-secondary school students. More studies should be conducted to investigate factors that may affect children’s learning behaviors and characters in digital-based language learning.

Figure 3. The Research Settings of the Empirical Studies

3.2 Research themes of game-based language learning

3.2.1 GBLL for linguistic knowledge and skills

Among the reviewed articles, we found eighteen studies focusing on the themes of GBLL for improving language learners’ linguistic knowledge and skills. As we can see from Figure 4, over half of these studies (9 articles) explored the potential benefits of digital games for enhancing language learners’ overall language proficiency, language performance or target language output. Others dealt with specific linguistic skills achieved through effective learning tasks in game-based language learning, such as listening, speaking, writing and vocabulary.

Figure 4. Research Themes of GBLL for Linguistic Knowledge and Skills

3.2.2 GBLL for non-linguistic knowledge and skills

Fourteen studies were identified as to explore the themes of digital games for improving learners’ non-linguistic knowledge and skills (see Figure 5). A variety of aspects concerning their non-linguistic knowledge and skills were discussed, among which, five studies centered on fostering language learners’ communicative skills, strategies or their social cultural interactions through the applications of
digital games. The benefits of digital games for other non-linguistic knowledge and skills also include boosting learners’ critical thinking, cross-cultural competence, and collaborative learning skills. Be more specific, Reinders and Wattana (2014) reported on a study into the effects of digital game play on learners’ communicative skills with a popular online role–playing game Ragnarok Online. Grantham et al. (2009) discussed how learners’ learning experience through the virtual reality games may impact their cultural knowledge acquisition, positive attitudes toward the target culture and their cross-cultural adaption or adjustment.

3.2.3 GBLL for language learner characteristics

Individual differences in second language learning are the main factors affecting second language learning. Among 35 reviewed studies, 12 articles focused on language learner characteristics (see Figure 6). The series of challenges presented by digital games are unpredictable and competitive, which can stimulate the curiosity of language learners and become an intrinsic motivation. This is why online games can serve as learning tools, providing learners with a brand new way of entertainment (Ampatzoglou & Chatzigeogiou, 2007). Six articles explored the impact of games on learner motivation. The benefits of digital games for other language learner characteristics also include promoting learner perceptions and attitudes, and learner engagement. This literature review indicated that existing studies have paid less attention to factors that affect students’ language learning, such as learner self-efficacy and language ability tendency.

3.3 Strengths and challenges of game-based language learning

3.3.1 Strengths of game-based language learning

Three advantages of using digital games to promote language learning are proposed based on the reviewed studies.

First, digital games have interesting user interfaces and game designs. Digital games constitute immersive virtual worlds which function as language-learning environments (Scholz, 2017). What’s more, digital games enhance learners’ learning interest and motivation (Jensen, 2017), and improve learners’ overall participation.

Second, users in digital games can be anonymous. As Grimshaw & Cardoso (2018) suggested that digital games with virtual environments tend to reduce anxiety for language learners and present a bridge to language use in the target language community.
Third, digital games provide real-time interaction. Immediate feedback from peers in the game can effectively improve learners’ engagement (Castañeda & Cho, 2016), and images and sound effects in human-computer interaction also play a positive role in encouraging users to learning continuously. Digital games seem to present a diverse and linguistically complex social-semiotic environment for L2 learners of English, and encourage learners’ collaborative learning skills and problem-solving ability.

3.3.2 Challenges of game-based language learning

Three challenges are summarized according to the reviewed studies.

First, students may have difficulty in applying this new technology to their language learning as digital games is different from traditional teaching practices. Reinders & Wattana (2015) found that a small number of participants in the study said they did not like gameplay and they hoped there would still be sufficient face-to-face instruction as well.

Second, language teachers need to be fully equipped with theories and trainings on digital game-based learning. Peterson (2012) showed that some sentences used in digital games may be grammatically incorrect, which would be a potential risk for learner participation in MMORPGs as well as in other types of CMC environment where access is provided to native speaker interlocutors. Therefore, teachers need to select appropriate digital games for students.

Third, it is expensive to develop suitable digital games for language learning, and there are high requirements for computers and Internet service, which many schools and language labs are not able to meet. This may make the popularity of digital game-based learning harder to achieve.

4. Conclusion

This study reviewed 35 studies on digital game-based language learning from 2009 to 2018. From the three aspects of the research trends, research themes and promises and challenges, a coding and analysis framework was constructed for a systematic literature review. It would be conducive to follow-up studies to use the content analysis method to conduct a systematic literature review of the relevant literature on the application of emerging technology education.

Through this review and analysis, a number of studies were carried out to ascertain the suitability of digital games in language learning. Digital games boast strong application potential in the field of language learning due to their interesting, anonymous and interactive features. However, the deep integration of digital games and language learning must rely on the cooperation in the field of language learning and educational technology, especially reasonable teaching design, well-targeted system development and long-term in-depth empirical research.

Acknowledgements

This research is supported by the Fundamental Research Funds for the Central Universities (2019XD-A04) and the Project of Discipline Innovation and Advancement (PODIA)-Foreign Language Education Studies at Beijing Foreign Studies University (2020SYLZDXM011).

References


The effect of a more knowledgeable other on resilience while playing single-player puzzle video games

Megan PUSEY*, Kok Wai WONG& Natasha RAPPA*
*College of Science, Health, Engineering and Education, Murdoch University, Australia
*megan.pusey@murdoch.edu.au

Abstract: Resilience refers to a person’s mental ability to adaptively deal with challenges in life. Video games (both commercial games and serious games) have been used as effective resilience interventions. There is some evidence that commercial puzzle video games could increase resilience as they involve overcoming frustration to succeed. This research explores if single-player commercial puzzle video games can be used as an effective intervention to improve resilience.

Participants were adolescents who attended an after-school club for 8-10 weeks. This paper presents the case studies of two club participants and their gameplay experiences. Data was collected through surveys, interviews, gameplay recordings and journals. In both case studies the participants both give and receive guidance and support from others. Having a ‘more knowledgeable other’ present while playing a challenging game helped participants deal with frustration and persevere. This paper provides a first step towards exploring the relationship between puzzle video game play, resilience and social support from others.

Keywords: resilience, video games, zone of proximal development, more knowledgeable other, social support

1. Introduction

Resilience is a person’s mental ability to adaptively deal with challenging situations (Masten, 2015; Pusey, Wong, & Rappa, 2020; Yeager & Dweck, 2012). Cassidy (2016) found resilience in the context of education/learning consists of three constructs - perseverance, adaptive help-seeking and the ability to deal with negative affect. Resilience is a trait that can be improved through interventions (Khanlou & Wray, 2014; Neill & Dias, 2001; Yeager & Dweck, 2012). Video games, both serious games and commercial off-the-shelf (COTS) games, have successfully been used as interventions to improve resilience across different contexts (Pusey et al., 2020). However, most of this research is aimed at either crisis response professions (e.g. paramedics) or people with existing conditions (e.g. depression) (Pusey et al., 2020). There is little research on using video games as a pre-emptive tool to increase resilience especially in the context of education.

There is some evidence that suggests COTS puzzle video games could be used as a pre-emptive tool to develop resilience (Scholten, Malmberg, Lobel, Engels, & Granic, 2016; Schoneveld et al., 2016; Tichon & Mavin, 2016). Schoneveld et al. (2016) suggests COTS puzzle games could train players in “resilience in the face of failure” and increase self-efficacy. This research seeks to intentionally explore the link between COTS puzzle video games and resilience.

Well-designed video games, especially puzzle games, are designed to be rich learning experiences allowing players “to operate at the outer edge of their regime of competence” (Gee, 2003, p. 70). While single-player games are designed for individuals to play alone, they are rarely completed in isolation. Gameplay is shaped “by people and material resources present in the room but invisible ‘in-game’” (Stevens, Satwicz, & McCarthy, 2008, p. 44). Social aspects of play are important to consider because they impact learning and resilience. In the Zone of Proximal Development (ZPD) a task is beyond a person’s ability to do alone and can only be completed with help from others (Vygotsky, 1980). A More Knowledgeable Other (MKO) provides guidance and encouragement to help a person complete a task in their ZPD. Neill and Dias (2001) found social support was a significant
factor in increased resilience in outdoor adventure education. In education, research has found cooperative learning, peer tutoring and social support can help increase resilience (Downey, 2008; Wilks, 2008). This research explores the role of the MKO in learning while playing single-player puzzle video games and how this relates to resilience.

There is little research on the role of a MKO in learning from video games. Luckin and Du Boulay (1999) used the serious video game EcoLab to help students learn about food webs in science. In a review of EcoLab research Luckin and du Boulay (2016) suggest “there is an increasingly important role for a More-Able Partner, whether in the form of software, people or some combination of the two” (p. 429). However, the EcoLab research focused on learning as the outcome, not resilience. Nardi, Ly, and Harris (2007) looked at learning conversations in World of Warcraft. They found “learners accomplish more with the aid of experienced peers than they could on their own” (Nardi et al., 2007, p. 8). The research found players offered each other positive encouragement, helped avoid frustration and gave a sense of moving forward. The research on EcoLab and World of Warcraft does not focus on resilience but hints that a MKO could impact the factors that contribute to resilience.

2. Research Aims

1. Does a MKO affect a person’s perseverance?
2. Does having a MKO in the room lead to more adaptive help-seeking?
3. Does a MKO help a person deal better with negative affect (frustration)?

3. Methodology

Participants were high school students (13 – 18 years old) who attended a weekly after-school club for one school term (8-10 weeks). The sessions involved playing a range of COTS single-player puzzle video games including The Witness, Baba Is You and Untitled Goose Game. This intervention was run over several weeks as longer interventions have been found to be more effective at increasing resilience with longer lasting effects (Khanlou & Wray, 2014; Yeager & Dweck, 2012). Venues included public libraries and community centres. This research was designed to capture “much better descriptions of what people actually do and learn playing video games under as naturally occurring conditions as possible” (Stevens et al., 2008, p. 42).

Single-player games were chosen to avoid competition and encourage participants to help each other. The games were chosen to be accessible to people who do not have much experience playing video games. All the games used are rated G by the Australian Classification Board.

Data was collected through surveys, journals, interviews and gameplay recordings. Gameplay recordings captured the participant’s laptop screen and voice. This exploratory research is presented as two case studies (Yin, 2014). Participant names have been replaced with pseudonyms.

4. Results

4.1 Case Study 1: Lena

Lena plays video games at home weekly, usually on console, smartphone, or tablet. Lena enjoys games such as Mario Kart and Minecraft. She had never played any of the games used at the club before. Lena would often comment on how difficult the puzzles were, muttering quietly to herself “this is so hard.” Additionally, she often celebrated her success by loudly exclaiming “I did it!” Lena frequently asked for help from either the first author or other participants in the room.

In the example seen in Table 1 Lena is attempting a sequence of puzzles where the player must figure out the ‘rule’ for black and white square symbols. The transcript shows the first author acting as a MKO for Lena who is struggling with the black/white square puzzles.

Table 1
Excerpt of transcript from gameplay recording of Lena

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Time</th>
<th>Transcript</th>
<th>Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lena</td>
<td>14:15</td>
<td>[whispers to herself] What?? (sounding frustrated)</td>
<td>Lena attempting black/white puzzle</td>
</tr>
<tr>
<td></td>
<td>14:18</td>
<td>How do you do this one?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14:20</td>
<td>[whispers to herself] This is so hard.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14:34</td>
<td>[to Researcher] I don’t know how to do this one this is so confusing.</td>
<td></td>
</tr>
<tr>
<td>Researcher</td>
<td>14:51</td>
<td>How’d you go?</td>
<td>Researcher walks over to Lena’s computer</td>
</tr>
<tr>
<td>Lena</td>
<td>14:52</td>
<td>I can’t do this one it’s really hard.</td>
<td></td>
</tr>
<tr>
<td>Researcher</td>
<td>14:54</td>
<td>So maybe the rule is not you have to go around all of them (the squares), maybe it’s something else.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15:06</td>
<td>Keep going.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15:19</td>
<td>Good job.</td>
<td>Lena solves black/white puzzle</td>
</tr>
</tbody>
</table>

Even though Lena had correctly solved four black/white square puzzles prior to this transcript she still did not understand the puzzle mechanic. The first author is careful not to correct Lena’s incorrect hypothesis of the rule, instead encouraging her to keep going and figure it out for herself. When Lena gets stuck on a puzzle she takes the initiative to ask for help. In later sessions Lena is often seen offering other participants support either by giving clues to puzzles she has solved or offering encouragement.

Table 2

**Lena’s journal entry for one of The Witness sessions**

| Session 2 | I felt frustrated at this game because alot of these puzzles were really hard and I needed some help with it. We played for an hour for about 4-5pm. It was fun but hard at the same time. When I needed help I got Megan to help me solve them. I would tell my friends about this game to see how well they can do and how far they can get in it. I enjoyed it alot. |

Lena’s journal entry in Table 2 show how frustrated she felt during the session but her choice to keep playing demonstrates her perseverance. Lena recognises when she needs help and readily asks for it from a MKO, showing adaptive help-seeking.

4.2 Case Study 2: Jessie

Jessie came to the club with Gabe, a friend from school, and his sister Ashe. Jessie plays video games at home weekly, usually on a computer or tablet. Jessie’s favourite video games are Minecraft, Roblox and Five Nights at Freddy’s.

In this example Jessie is trying to solve a difficult puzzle involving two mechanics, the black/white squares (seen in Case Study 1) and the stars. Following the transcript in Table 3 Jessie spends the rest of the session (another 20 mins) attempting this same puzzle.

Table 3

Excerpt of transcript from gameplay recording of Jessie

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Time</th>
<th>Transcript</th>
<th>Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jessie</td>
<td>11:40</td>
<td>This is hard. Why did life have to give me lemons?</td>
<td>Jessie sitting next to Gabe.</td>
</tr>
<tr>
<td></td>
<td>11:53</td>
<td>I dunno does this work? No, I need help.</td>
<td>Submits incorrect solution.</td>
</tr>
<tr>
<td></td>
<td>12:04</td>
<td>What if I just went straight through the middle?</td>
<td>Submits incorrect solution, all symbols flash red.</td>
</tr>
</tbody>
</table>
Jessie shows adaptive help-seeking as he first asks Gabe for help. Unfortunately, Gabe has not done this section of the game (so is not a MKO) and tells Jessie he cannot help. Jessie then asks the first author for help. This gameplay recording highlights Jessie’s perseverance as he spends 25 minutes trying to solve one puzzle without giving up. The first author offered guidance and support throughout the session but did not tell Jessie the answer. Jessie’s self-talk at the beginning of the transcript shows how frustrated he is at not being able to solve the puzzle. Gabe and Jessie both admit The Witness makes them feel frustrated, yet they still want to play. Being able to share his frustration with someone who understands, even though they are not a MKO, seems to help Jessie deal with his frustration. Why are Jessie and Gabe motivated to persevere? Perhaps, because there is a MKO in the room to help them. Or, maybe it is the feeling of achievement after solving a difficult puzzle, mentioned in the interview in Table 4.

Table 4

| Researcher: How did you feel when you got really stuck on one of the puzzles in The Witness? |
| Jessie: Quite annoyed, irritated. |
| Ashe: Frustrated as well. |
| Researcher: And how did it feel when you finally got the answer to that puzzle? |
| Jessie: Quite relieving. |
| Ashe: Yeah it made you feel motivated to complete more puzzles. |
| Jessie: Or disappointed if you find out you have to do another harder one [laughs]. |
| Researcher: If you got really stuck what were some of the things you did to help you solve that really frustrating puzzle? |
| Jessie: Well you could go ask someone else for help like Gabe who was there (at the club). Or maybe you. Or probably just go to another puzzle somewhere else. |
| Ashe: Or look around for clues around the area that you’re in to help you figure out the puzzle. |
| Researcher: Why would you give someone a clue and not just tell them the answer for this game? |
| Jessie: So instead of just them not using any brainpower they use some at least. |
| Ashe: And like if you’re telling them the answers then they struggle to figure out other puzzles because they haven’t had to use their brain they’ve just been told the answer, so they just expect the answer all the time. |

Jessie and Ashe mention wanting to understand puzzles instead of just being told the answer. In the survey Jessie mentions his favourite game from the club was The Witness writing, “it felt more fulfilling when I completed a challenge or puzzle in The Witness.”

The last few sessions of the club were interrupted by COVID-19 restrictions meaning Jessie played Untitled Goose Game at home. In the interview Jessie mentions quitting Untitled Goose Game before finishing the first area as he struggled with the controls saying “it was quite annoying trying to get the keys off the gardener. It was a little hard to sort of manoeuvre the little goose around.” He did not
have a MKO available to him at home. In comparison, all participants who played *Untitled Goose Game* during the club got past this first area. Perhaps, with support from a MKO he could have overcome this difficulty.

5. Discussion

Having other people physically present in the same room as the player makes a big difference to the gameplay experience, even with single-player games, mirroring findings from Stevens et al. (2008). As seen in the above case studies participants regularly asked a MKO for help when they were stuck. Games do offer good feedback as it is just-in-time and instant. However, it is possible for players to misinterpret this feedback. When acting as a MKO the first author tried not to tell participants the puzzle mechanics but to give hints based on the players existing level of understanding. The benefit of a MKO is their ability to offer personalised feedback and support based on the current level of understanding of the player.

Jessie and Lena were both recorded regularly persevering on difficult puzzles for up to 10 mins before asking for help. In comparison easier puzzles took between 20 secs – 4 mins. Does this mean regardless of if a MKO is present people will persevere? Not necessarily. Jessie showed great perseverance in *The Witness* but quit *Untitled Goose Game* in the first area even though the puzzles in *Untold Goose Game* are easier. The difference was that Jessie played *Untitled Goose Game* exclusively at home (due to COVID-19 restrictions) without access to a MKO. All other participants who played *Untitled Goose Game* at the club (with access to a MKO) got past the first area. In Case Study 2 social support and empathy from a friend seemed to help Jessie persevere, spending up to 25 minutes on one puzzle. This aligns with existing literature on the positive impact of social support on resilience (Downey, 2008; Neill & Dias, 2001; Wilks, 2008). The presented case studies suggest that social support and just-in-time personalised support from a MKO could help players persevere for longer before quitting. The key difference between a friend and a MKO is the type of support they can offer. Someone who is not a MKO can only offer encouragement. But, a MKO can offer encouragement and guidance on what to do next. If a player is on the verge of giving up a MKO can offer a hint and guide the player to success in order to motivate them to continue playing.

The gameplay recordings show that Lena and Jessie would try a few different strategies to solve puzzles before asking for help. Common MKOs included friends or siblings who had played game already or the first author. Surprisingly, the surveys and interviews indicated online sources such as YouTube were not commonly accessed for help. Perhaps it is the MKOs ability to offer personalised support that makes them the preferred choice compared to online sources. Lena, Jessie and Ashe all mention wanting to understand the puzzles, not just know the answers. Alternatively, it could be the existing relationship between the MKO and player that makes them the first point of contact for help.

Lena, Jessie and Ashe all mention finding the puzzles challenging and feeling frustrated yet they all continue playing the games. The participants also recognise there is a feeling of satisfaction after solving a puzzle because of the level of challenge. Luckin and du Boulay (2016) mention that EcoLab could not avoid students feeling frustrated, but perhaps avoiding frustration is not necessary. One possibility is that it is important to help learners overcome frustration and feel the rewarding satisfaction of completing a challenge. Even sharing the frustration with a friend who understands, seemed to help Jessie deal with his frustration. Maybe it is just a sense of social support, rather than a MKO, that is needed to help people deal with negative affect such as frustration.

6. Conclusions, Limitations and Future Work

This exploratory study investigated the role of a MKO in developing resilience while playing single-player puzzle COTS video games. Social support from a friend (not necessarily a MKO) helped improve perseverance and dealing with negative affect. A MKO was able to offer guidance as well as encouragement. Well timed guidance from a MKO can help players on the verge of quitting experience success and feel motivated to continue playing. These findings suggest that having a MKO present in the room can help improve a player’s resilience. The results match existing literature showing that
social support increases resilience and that the social aspect of games makes a big difference to the gameplay experience (Downey, 2008; Neill & Dias, 2001; Stevens et al., 2008; Wilks, 2008). The data collected is rich and goes some way to understanding the thought processes of players as they solve difficult puzzles. However, there is only a small number of participants. Future research into resilience and gameplay cannot focus only on individuals but also needs to incorporate the social and in-room context. Designers of single-player games need to consider the social context when creating games, particularly serious educational games.

Acknowledgements

This research has been supported by the Australian Government Research Training Program. We would also like to thank Thekla Inc for providing copies of *The Witness* and Arvi Teikari for providing discounted copies of *Baba Is You* for this research.

References


VRetorik: A VIRTUAL REALITY VIDEO GAME TO IMPROVE PUBLIC SPEAKING SKILLS

Enrique FUERTES, Javier NAVARRO, Jagoba MONTES, Manuel GONZALEZ, Borja MANERO
Complutense University of Madrid, SPAIN
enrifuer@ucm.es
javina@ucm.es
jagomont@ucm.es
manuel.gonzalez@ucm.es
bmanero@ucm.es

Abstract: A Google search for "fear of public speaking" is enough to get more than eight million results. This is no coincidence since 75% of the population (glossophobia, 2011) suffers from this phobia. This is why there are numerous studies carried out over the last few years (Anke W. Blöte, 2009). In addition to these, some tools seek to help improve the ability to speak in public. This is the case of Orai (Orai, 2018) or Chiara (Chiara, n.d.), who can create personalized training after recording and analyzing a speech. Not everything is focused on technology and training, you can also find board games, as is the case of Retorik (Juanjo Mestre y Marta Segarra, n.d.). Retorik is a card game in which players must improvise a speech, improving their communication skills by playing. Both games and technology can help improve public speaking skills. In this project, both worlds have been combined to create a virtual reality video game capable of analyzing the speaker and giving him/her feedback to improve his/her communication skills. This paper details the design and creation of VRetorik, an educational virtual reality video game aimed at improving these skills. It relies on improvisation and creativity to gamify the player's experience. Using artificial intelligence systems, the content of the speech, and the emotions transmitted by a speaker is analyzed to provide feedback. In this way, it is intended that the speaker improves the ability to speak in public by playing.

Keywords: public speaking, educational video game, virtual reality, speech, gamification.

1. Introduction

Throughout the human academic period, there are hardly any mechanisms to improve public speaking skills (Ott, 1998), which is why fears and phobias arise when facing an audience to give a speech.

To address these fears, some therapies provide a coach to improve public speaking skills, who offers a personalized plan to reduce the fear of public speaking (Parker, 2003).

Thanks to technological advances, tools have been developed in which virtual reality plays an important role. These tools are designed to improve public speaking skills (Cristina Botella, Javier Fernández-Alvarez, Verónica Guillén, 2017). Virtual Reality allows us to generate a safe virtual environment, as realistic as possible to the situation to be represented. An example is BeFearless (Samsung, 2016), a tool developed by Samsung.

However, many of these tools simply generate a virtual environment in which to train but do not provide real-time feedback on how the user is doing it and do not work on a key aspect such as improvisation.

On the other hand, several studies [ref] show how important and effective video games are when it comes to improving skills in fields such as mathematics (Devlin, 2011) and programming such as CodinGame (codingame, 2016), among others. Besides, virtual reality is starting to be used in language learning (Klaus Schwienhorst, 2002).

Apart from technological tools and training methods, there are board games focused on improving communication skills. A relevant case is Retorik, a card game with which through creativity,
improvisation, and entertainment, its players can gain confidence and skills to lose their fear of public speaking.

In this paper, we present the design and implementation of VRotorik, an educational video game (based on Retorik) that focuses on improving public speaking skills through improvisation.

The document is structured in such a way that it first reflects the objectives of the project. Then, it focuses on how the transformation from the board game Retorik to the video game VRotorik has been carried out in virtual reality (see Figure 1). Next, the dynamics of gamification that have been followed will be discussed and finally, the conclusions and future work of the project will be detailed.

2. Target

The main objective of this project is to develop an educational video game in virtual reality, which is capable of providing real-time feedback to the player's presentation. In this way, we seek to improve their public speaking skills through improvisation.

At the same time, the aim is to create a virtual space where the user feels comfortable and where he can rehearse a speech as many times as he needs to in a realistic environment. You will also be able to familiarize yourself with the feelings of discomfort or oppression generated by this type of phobia and master them little by little.

To do this possible, a gaming environment will be developed that is motivating for the player. This has to be a safe environment in which the player is not afraid to make mistakes. Finally, it is important, as in any video game, to balance the level of difficulty so that it is challenging but not frustrating for the player.

Finally, useful feedback must be offered to the user based on the content of the speech and the emotions it conveys. The player must know their successes and errors to improve.

3. Game Design

This section first explains the dynamics of the board game on which VRotorik is based. Next, the process of adaptation and improvement of the original board game is detailed, and finally, the gamification process is used.

3.1. Retorik: the board game

Retorik is an educational and creative project, funded by crowdfunding. It aims to improve communication skills when speaking in public. It focuses primarily on improving the player's ability to improvise, a key aspect of public speaking.

The game begins by selecting a card at random from the theme deck and five others that include random phrases. The player has to talk about the topic that has come up by introducing the 5 random phrases into their speech for 1 minute. At the end of the time, the rest of the players must evaluate the speech, determining whether the topic has been satisfactorily addressed and whether they have introduced the phrases. If so, the player keeps the theme card and the one who has more theme cards at the end of the game wins (see Figure 1).
3.2. From Retorik to VRetorik

VRetorik is divided into two environments (see Fig. 2). On the one hand, the virtual environment (VE), where the player must present his speech in front of an audience following the rules of the game Retorik. On the other hand, there is the analysis environment (EA), where all the data collected from the speech is processed in real-time, and feedback is generated to the player about his speech.

3.2.1 Aleatory sentences

It is established that a key aspect of public speaking is the ability to improvise. That's why the dynamics of the game is focused on getting random phrases into the speech most creatively and coherently.

To simulate the selection of five cards as in the original game, five random phrases will be generated at the beginning of the game from a pre-established list. A real-time voice transcript is then made so that when the player says the sentence, the system recognizes it and removes it from the environment (see Fig. 3). The randomly generated sentences must be the same as those recited by the player, just as in the original game. Also, the speech transcript is stored for later analysis and processing.
3.2.2 Topic

VRetorik implements a mechanism for identifying topics from the audio transcript. To do so, at the beginning of each game, a random theme is generated on which the player must deal with the speech that is going to be articulated. That topic is offered to the player and he must speak about it for one minute.

Along the process of identifying the theme, our algorithm is supported by the Grafeno tool. This tool allows us to create a concept network from a plain text, with a series of transformers and operations, clusters of concepts are generated from the original text. The clusters that have a higher degree of connection are selected, this value is set to discard less connected clusters, and those with a higher degree are stored next to the concept to which they refer.

Once, an algorithm has been implemented from the highest degree clusters generated in the concept network, the three concepts of the highest degree clusters are extracted. With the use of natural language processors (NLP) such as Wordnet and NLTK, a process of searching for synonyms of the concept in the original tokenized text is achieved to identify the topic. For each occurrence of the concept synonym in the tokenized text, more weight is given to the topic being processed. In the end, the topic that has obtained the most weight after this screening is selected and identified as the topic of the speech. Finally, this topic is compared with the one originally requested by the user to decide if the player has adjusted to the requested topic.
3.2.3 Sentiment analysis

Although the original game does not ask for anything in terms of emotions in the player's speech, it has been decided to use the technological advantages to improve the gameplay. To do this, during the game, the player is asked to recite each of the phrases with the feeling associated with it generated randomly.

Once the game has started, the system uses a voice feeling detection algorithm. The use of an algorithm that detects feelings from the audio is much more accurate than the algorithms that detect them in the text, such as TextBlob. For this analysis, the system detects the moment in which the phrase is recited and stores the audio fragment for analysis. This step is repeated for each of the five phrases that the player must recite. Finally, five feelings associated with each phase of the game are identified: happy, neutral, sad, angry, and fear.

For the analysis of feelings by voice, Vokaturi has been used, which, based on the audio recorded during the game, detects the feeling expressed in the game. According to the tests carried out from audios recorded by actors and actresses, Vokaturi gives us 66.5% accuracy in detecting feelings.

3.2.4 Public

Finally, and to make the game more dynamic, a virtual audience has been designed to react to the evolution of the paper (see Fig. 5). Depending on the development of the presentation, the listeners interact with the speaker. These actions are managed through the data obtained in real-time from the game and each character in the audience reacts differently thanks to the design of a state diagram.

![Fig 5. Public clapping and shouting during the talk](image)

3.3. VRetorik Gamification Dynamics

3.3.1 Score

At VRetorik, a scoring mechanism based on three factors has been implemented:

- Random phrases: for each phrase that the player enters in his speech, he will be given a score of 5 points for each one, up to a maximum of 25 points.

- Feelings: if the player can reproduce the feeling requested in each phrase, the score will be increased by 5 points, up to a maximum of 25 points.

- Theme: After the speech, if the theme identified in the speech by the implemented tool coincides with the theme proposed at the beginning of the game, a score of 50 points will be awarded.

With the three factors mentioned above, the player obtains a maximum score of 100 points. Once the count of random phrases introduced in the speech and feelings expressed in them is finished, they are processed to analyse how many of them have been successfully overcome.

The scoring algorithm is based on the following formula:

\[
((5 \times N^r \text{ right sentences}) + (5 \times N^r \text{ right sentiments})) + 50 \times [0 \mid 1 \text{ if topic is wrong} \mid \text{correct}]
\]
4. Conclusions and future work

An educational video game has been developed, which through improvisation and creativity improves public speaking skills. To this end, the dynamics of Retorik have been maintained in a gamified virtual reality environment. The video game is capable of randomly generating the phrases that the player must include in his speech, making a voice transcription of the player's speech to subsequently identify the topic of the speech, and, finally, from the recorded audio of the speech, perform a sentimental analysis. The sentimental analysis makes the improvisation much stronger in the user because he will have to make an effort to capture the feeling given at the beginning of the game.

For the identification of the topic, a method of regressive derivation was realized from the topics obtained by Grafeno. This process consists of locating the root of the word by eliminating its affixes. Following this identification system, the expected accuracy was not achieved. That is why a synonym search algorithm was chosen from the extracted themes. After this process, their appearances in the original text are counted to add more relevance to the concept from which these synonyms are derived.

As future work, it is proposed to implement a more complex punctuation algorithm. We also want to develop an algorithm that adapts the feeling that the player must interpret each random phrase according to its morphology. We are currently working to improve the video game and provide it with greater syntactic flexibility.

Due to the state of the Covid-19 pandemic, the desired experiments could not be carried out with users in schools. Experiments are planned to test the effectiveness of the video game and its scope.

Acknowledgments

To sum up, I would like to thank our project manager Borja Manero Iglesias and co-project manager Manuel Gonzalez Rio for their support.

References

Using a Situated Speech-based Holographic Projection System to Learn the Analects of Confucius

Yang-Hsin Fan\textsuperscript{a*}, Chi-Fang Huang\textsuperscript{b} & Zhi-Hong Chen\textsuperscript{a,\&b}

\textsuperscript{a}Program of Learning Sciences, National Taiwan Normal University, Taiwan
\textsuperscript{b}Graduate Institute of Information and Computer Education, National Taiwan Normal University, Taiwan

*\texttt{benson8610124@gmail.com}

Abstract: In this paper, based on the concept of situated learning, we describe the framework of speech-based holographic projection, which is comprised of the elements of learning content, holographic projection as authentic context, and interactive activities. Based on the speech-based holographic projection framework, we applied the system to convey the content of the Analects of Confucius.

Keywords: Holographic projection, Speech recognition, Situated learning, The Analects of Confucius

1. Introduction

Confucius's teaching and reflections on learning, community, and the conduct of life have a profound influence on both Chinese society and the wider global community (McLeod, 2014). However, due to the Analects of Confucius is written by Classic Chinese, many students fail to understand its abstract thoughts by using the textbook and consider it as a big challenge to study it (Sung, Hwang, Chen, & Liu, 2019). Therefore, this study aimed to provide a speech-based holographic projection system to help students learning the Analects of Confucius.

2. Design framework

As the characteristic of situated learning in game-based emphasis on how learners participate in situated authentic context (Zhang, & Shang, 2016), Key tenets below are essential to support the construction of situated learning environment. In the paper of Green, Eady, and Andersen (2018), constructive learning approach, authentic context, and social interaction are the three main elements of situated learning. To foster the situated learning design in multimedia, this study proposes a speech-based holographic projection framework, which corresponds to the key tenets of situated learning theory claimed by Green et al. (2018) The core element of the framework includes content, authentic context, engaging interaction as illustrated in Figure 1.

![Figure 1. Speech-based holographic projection framework](https://example.com/figure1.png)
3. Implemented system

Based on the speech-based holographic projection framework, we applied the system to convey the content of the Analects of Confucius (as shown in figure 2). This system is made with Unity to build the scene, and Google Cloud kits to construct the speech recognition system and exported into APK file.

![Figure 2: The holographic projection](image)

The content of the system is excerpted from the Analects of Confucius. Three important stages of stories about Confucius are selected, and divide each into three points to express further information. The first stage emphasis on the background of Confucius. As the family situation and the interest to Rites of Zhou made him study harder, these experiences settle the basis of his achievement. In the second stage, the educational philosophy is the core. Introduced from how Confucius realized that education might assist the civilian, to his remarkable attainment in the education sector. Finally, this stage reviewed the achievement of Confucius on publication works, influence on the history at that time and to posterity. According to these narrative texts, the philosophy of Confucius and the influence on the posterity are directions to do critical thinking. Besides, objective knowledge is not only applicable to Chinese courses but also possible to impacts points of view in life.

The scenario consists of 3D models of Confucius, his students, and the background build up the authentic context which is helpful for learners to experience. Furthermore, 3D models of Confucius is also voiced by a real human, which makes the projected-Confucius closer to a real person. The authentic context brings the scenes thousands of years ago into reality and might be one of the most applicable situations that learners immerse in.

To interact with the 3D Confucius, the speech recognition system makes it possible for learners to learn by communicating with him. When awake the system through specific sentences invoice, which means there is no need to press button or type orders, 3D scenes about the keyword related to Confucius will show up in the hologram environment. With the speech recognizer, students might be able to pay attention to the learning target and can gain knowledge in one of the most straightforward ways. Also, there would be audio clips of the story from the perspective of both the narrator and Confucius, is played coordinate with the scene showed up in the environment. Through both ways of verbal and visual aids, plus the bi-directional activity of learners’ voice and the audio clips, we expected that learners would gain learning motivation and acquire the objective better.

As this study emphasizes the framework of speech-based holographic projection, we aim to solve some of the difficulties using AR and VR. The authentic context of the scenario and the interactivity of the speech reaction system plays an important role in this study.

Reference
Improving EFL Students’ Learning Achievements and Behaviors using a Learning Analytics-based e-book System

Mei-Rong Alice CHEN*, Rwitajit MAJUMDARb, Gwo-Jen HWANGc, Yi-Hsuan LINc, Gökhan AKÇAPINARD, Brendan FLANAGANb, & Hiroaki OGATAb

*Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taiwan
bAcademic Center for Computing and Media Studies, Kyoto University, Japan
cDepartment and Graduate Institute of English Language & Literature, Chinese Culture University, Taiwan
dDepartment of Computer Education & Instructional Technology, Hacettepe University, Turkey

Abstract: An e-book system allows students to access learning materials regardless of location or time. However, reading e-books without additional guidance could be inefficient for students and thereby affecting their learning engagement and learning outcomes. In this study, a learning analytics-based (LA) e-learning approach was proposed to resolve the situation. Moreover, experimental design methods were used to explore the relationship between the learning engagement, behavior, and achievement of students who learned with the LA e-learning approach. The research results show that the learning performance of high-engagement participants using the LA e-learning approach is better than that of low-engagement participants. In addition, the findings indicated that the proposed method could support low and medium learners to improve their learning achievement. These findings could be a valuable reference for those who intend to develop an effective learning analytics-based learning approach through e-book systems.

Keywords: E-book, e-learning, learning behavior, learning engagement, learning analytics

1. Introduction

Conventionally, English as a Foreign Language (EFL) learners passively receive the instruction given by teachers in the class without interaction, so they have little chance to engage in reading by themselves (Sanprasert, 2010). In addition, EFL learners generally rely on their teacher’s translation of the texts (Mollaei, Taghinezhad, & Sadighi, 2017); therefore, they tend to become disengaged from learning (Larsen-Freeman, 2000). Besides, high school English courses in many countries are mainly lecture-based and involve a large number of students. This arrangement hinders students from engaging with English classes, as instructors cannot monitor the students’ learning when the class is big.

When considering the use of technologies, it is important to think about whether they facilitate students’ engagement. Language learning with digital technologies offers EFL learners more chances to improve their engagement and can lead them to become autonomous learners (Yang, 2012; Zulkepli et al., 2018). Particularly, e-books have been shown to have a positive impact on enhancing literacy skills (Shamir, Segal-Drori, & Goren, 2018). Researchers have studied ebook systems and indicated that they could facilitate learners’ interaction with the learning content, such as when using an interactive e-book-based flipped learning approach (Hwang et al., 2017) to enhance both in-class and out-of-class learning. Their study showed that the learners had more chances to engage not only in class but also after class. In addition, previous studies have also suggested that learning logs from the learning system can assist students and teachers (Hwang et al., 2017). In addition, Hsieh and Huang (2019) claimed that while students benefit from the ebook features such as color markers and annotation, students’ learning
behaviors while using books have not been evaluated extensively. Little attention has been paid to visualizing and analyzing the relationship between learners’ strategies and their learning outcomes in the learning logs. Besides, few studies have looked specifically at learners’ engagement levels concerning ebook learning. Therefore, it is necessary to explore this aspect of learning analytics. This study investigated students’ engagement levels with the proposed learning analytics-based approach and their relationship with the students’ learning achievement and behaviors.

Therefore, in this study, a learning analytics-based (LA) e-learning approach was proposed to empower students to have more autonomous learning during the ebook reading process. An experiment was conducted with first-grade senior high school students in Japan. This study aimed to investigate the participants’ learning achievement, engagement, and learning behaviors. The following research questions guided the study:

RQ1. What are the impacts of the LA e-learning approach in terms of learning achievements of the high-engagement group and the low-engagement group using the LA e-learning approach?
RQ2. What are the impacts of the LA e-learning approach in terms of learning achievements of the low achievers using the LA e-learning approach?
RQ3. Is there any correlation between their learning behaviors and learning achievement?

2. Literature review

2.1 Learning engagement

Learning engagement has been defined as consisting of endeavors to learn (Strati, Schmidt, & Maier, 2017). Also, Han and Hyland (2015) stated that engagement means using recognizable strategies to support students’ learning outcomes. Chen, Hwang, and Chang (2019) claimed the significance of highlighting student involvement in autonomous learning engagement in learning systems. For example, Chen, Lu, and Lian (2019) compared learners’ engagement when playing a game, watching a video, and in the classroom. Previous studies relied on eye trackers (Kao, Chiang, & Foulsham, 2019), questionnaires (Abubaker & Lu, 2017), or coding (Abdelhalim, 2017; Li, Ma, Wang, Lan, & Dai, 2019) to evaluate students’ engagement. For example, Li and his colleagues (2019) studied pre-school children’s engagement by videotaping then coding their behavioral patterns and performances.

Other studies have found that engagement with the textbook was related to student learning achievement (Junco & Clem, 2015; Kuh, 2009), and the time and effort students devote to activities that are empirically linked to desired outcomes. Specifically, they identified that time spent reading tends to be one of the critical variables to predict students’ academic outcomes. Therefore, engagement in this study was explored in terms of the time students invested in their learning using the ebook system to achieve the expected learning outcomes. Their learning log data were generated from the ebook system to represent their learning engagement.

2.2 Learning approach and e-books in English language learning

The SQ3R reading approach is a famous reading comprehension method developed by Robinson (1946). It focuses on learners’ knowledge and comprehension, which are categorized as lower-level thinking items. The purposes of these six steps are to keep students engaged in reading when using the ebook system. The first two, scanning and skimming, can support learners in developing their reading comprehension (Anderson, Rourke, Garrison, & Archer, 2001), and facilitate their reconstruction of the text structure. Scanning requires readers to read rapidly to locate details-specific information (Brantmeier, 2002; Liaw, 2017) to get answers from the questions in the assignment or exam, while skimming is reading quickly to get a general overview of the paragraph. Question posing involves comprehending, constructing knowledge, and thinking about the questions during the learning process (Sung, Hwang, & Chang, 2016). Replying to the question helps promote students’ deep comprehension of the reading content, and enables them to think, address, and solve questions (Ye, Chang, & Lai, 2019). Learning engagement provides learners with a structured opportunity to scrutinize their learning (Verpoorten, Westera, & Specht, 2012).
3. Method

In this study, the impacts of learning analytics-based (LA) e-learning approach on students’ learning achievement, learning engagement, and learning behaviors were investigated. An experiment was conducted to capitalize on the quantitative method.

3.1 LA e-learning approach

In this study, an e-book reader, BookRoll, (Flanagan & Ogata, 2018), was adopted. The e-book reader provides features such as functions of the marker, quiz, and memo. Robinson's SQ3R was tailored as the LA e-learning approach and integrated into the e-book reader. In this study, question reply and reflection were added to the learning approach because they are categorized as higher-level thinking items (Pena-Shaff & Altman, 2015). Therefore, the revised learning strategies of Scanning Skimming, Question, Read, Reply, and Reflection were developed as the learning analytics-based (LA) e-learning approach for this study. The learners can use the proposed approach (i.e., Scan & Skim, Read, Reply, and Reflect) procedure to go through the learning in the e-learning system step by step, shown in Figure 1. This means that the participants first scanned for the main ideas and used the red marker function to highlight important ideas. At the same time, they skimmed for the general ideas, and they used the yellow marker function if they did not understand the gist of the paragraph. In the quiz function, they could read the questions that the teacher had prepared, and they read the text to find the answer and then reply to the question. The memo feature was to indicate sections where students could make their annotations at the page level. The last step of the LA e-learning approach was to use the memo function to record their reflection of what they had learned.

In addition, teachers can set questions in the Quiz function. Students were required to answer the questions in the quiz function after scanning and skimming. When they had done the scanning and skimming and had replied to the question, they should then write their reflection of the reading in the memo function. In the LA-learning system, the analysis tool displays the learning logs from the e-book reader, including the red markers for scanning, the yellow markers for skimming in the Marker list panel, and the quiz function for the question reply scores in Quiz Score panel, and the memo reflection in the memo panel. The learning analytics-based (LA) e-learning activities were recorded by the analysis tool as students’ learning log data, as shown in Figure 2.
3.2 Participants

This study used an experimental design combined with a quantitative method. One of the Integrated English Competency (IEC) classes of students were recruited from the first-grade high school in the west-central Honshu Island, Japan. The IEC course aims to assist students in vocabulary acquisition, grammar use, and reading comprehension skill development. The participants (n=40, 17 males, 23 females) adopted the LA e-learning approach, and their ages ranged from 15 to 16. At the time of the research, they had been studying English for approximately four years. Their English proficiency level is roughly equal to a pre-intermediate to intermediate level or a B1 in the CEFR scale. Their English classes comprised two 50-minute lessons a week throughout the academic year. One specific textbook-NEO See-the Modern Approach (Watanabe, 2016) was assigned and approved for the IEC course. The textbook is a collection of Basic English readings selected from the themes that frequently appeared in recent university entrance examinations in Japan.

3.3 Experiment procedure

An experiment was conducted on the Integrated English Competency (IEC) course in the fall 2019 term to evaluate the proposed learning approach on the learning achievement, and learning behavior of the students learning. The participants in the LA e-learning approach were instructed to appropriately operate the tablet and the e-learning system to learn the content. The teacher started a unit by encouraging the students with warm-up activities. Keywords were presented and reinforced through sentence-building exercises, explanations, and exercises on grammar rules. The purposes of the teacher’s instruction were to activate students’ background knowledge by asking general to specific questions through modeling, guiding the students to help them become familiar with the vocabulary, and monitoring their comprehension by asking questions from the textbook.

The experiment was conducted on the two units of an Integrated English Competency course, which aims to enhance the high school students’ English reading skills. Before the experiment, the participants spent one week on the pre-test to evaluate their prior knowledge of English reading. Following that, the participants were required to learn and use the LA e-learning approach in the e-book reader for four weeks. After four weeks of the learning proposed approach, the participants took the post-test regarding the unit content they had learned during these four weeks. In this study, students were learning Units 11 and 12.

3.4 Data analysis

One English pre-test and post-test were created, and the test validity was ensured by the high school teachers from the high school and the researchers. The pre- and post-test as the learning achievement tests in this study consisted of three sections focusing on receptive English reading skills. The pre-test measured the participants’ English language proficiency, and the post-test was to monitor the students’ learning achievement for the two learning units. The questions were from Units 11 and 12 in the textbook, and the test comprised 27 multiple-choice questions. There were three sections for learning.
achievement, including lexical, semantic, and reading comprehension questions, with a total score of 100. The Cronbach’s alpha value of the achievement test is .97.

To examine the effect of the LA e-learning approach on English learning achievement to analyze the transition patterns of the students’ achievements across the pre- and post-test, the interactive Stratified Attribute Tracking (iSAT) method was used (Majumdar & Iyer, 2016). iSAT can visualize the distribution of the pre-test and post-test scores of the cohort of learners and track the changes in the overall learning results. The scores of both the pre- and post-test were stratified based on the top 25% (10 participants) as high proficiency, the middle 50% (10 participants) as medium proficiency, and the bottom 25% (10 participants) as low proficiency (Papi & Abdollahzadeh, 2012).

Learning behaviors include annotations using the red marker, yellow marker, memo, and attempting quiz functions of the e-book reader. Such learning behaviors were automatically logged within the system. The counts of the specific learning behaviors were accessible to both the researchers and teachers through the learning dashboard called Analysis Tool. The participants’ learning behaviors were extracted from the number of times the red marker highlight was used for scanning, the number of times the yellow marker was used for skimming, the quiz scores as replying to questions, and the number of times memos were made as to their reflection of the reading. The participants’ time spent learning the unit was analyzed as to their learning engagement. Thus, the participants’ levels of engagement and learning behaviors were collected directly from the learning logs of the e-learning system during the 4 weeks of study. Then the relationships between the learning achievement, engagement, and behaviors were computed using the statistical analysis system, SPSS.

Besides, the differences between the high- and low-engagement EFL learners learning the two units of the two-class sections were analyzed. The participants were ordered from the highest to the lowest engagement groups based on their level of engaged behavior obtained through the e-learning system. The learning logs in the e-learning system of participants’ time spent learning the materials were analyzed as to their learning engagement. Engagement groups were based on the total time spent learning, as recorded in the learning logs obtained through the e-learning system. The grouping was calculated as the top 25% (10 participants), the middle 50% (20 participants), and the bottom 25% (10 participants) groups accordingly (Papi & Abdollahzadeh, 2012), and the cohorts were labeled as the high-engagement, moderate-engagement, and low-engagement groups. According to the ANOVA results for the specified groups, it did truly represent the different engagement groups (F = 90.33, p < .05, η² = 0.17). The results of the analysis strongly confirmed a significant difference in the two units’ average time (in minutes) of the high-engagement group (M = 51.60, SD = 4.72), the moderate-engagement group (M = 35.28, SD = 5.43), and the low-engagement group (M = 22.55, SD = 3.49), as displayed in Table 1. Furthermore, post hoc analysis was performed to examine specific differences in the engagement of the three groups.

Table 1. ANOVA result of the learning time of the three levels of engagement

<table>
<thead>
<tr>
<th>Level of Engagement</th>
<th>N</th>
<th>M(min)</th>
<th>S.D.</th>
<th>F</th>
<th>Post hoc tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-engagement (a)</td>
<td>10</td>
<td>51.60</td>
<td>4.72</td>
<td>90.33*</td>
<td>a &gt; b</td>
</tr>
<tr>
<td>Moderate-engagement (b)</td>
<td>20</td>
<td>35.28</td>
<td>5.43</td>
<td></td>
<td>b &gt; c</td>
</tr>
<tr>
<td>Low-engagement (c)</td>
<td>10</td>
<td>22.55</td>
<td>3.49</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *p<.05

4. Results

4.1 Analysis of English learning achievement

In order to answer the first research question concerning the differences between high-engagement and low-engagement EFL learners in terms of the effect of the LA e-learning approach on English learning achievement, the participants were ordered from the highest to the lowest engagement groups based on their level of engaged behavior obtained through the e-book reader. The top one fourth (10 participants), the one half (20 participants), and the bottom one fourth (10 participants) groups were
distinguished and labeled the high-engagement, moderate-engagement, and low-engagement groups, respectively.

The ANOVA outcome of the comparison of the three levels of engagement for the LA e-learning approach is shown in Table 2. The result indicated that there was a significant difference in the mean scores of the learning achievement test for the high-engagement group (M = 90.70, SD = 2.50) and the low-engagement group (M = 65.10, SD = 6.88), F = 39.58, p < .001, as presented in Table 3. In other words, high-engagement in the LA e-learning approach can successfully promote students’ learning achievements.

<table>
<thead>
<tr>
<th>Learning achievement</th>
<th>N</th>
<th>M</th>
<th>S.D.</th>
<th>F</th>
<th>Post hoc tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-engagement (a)</td>
<td>10</td>
<td>90.70</td>
<td>2.50</td>
<td>39.58**</td>
<td>a &gt; b</td>
</tr>
<tr>
<td>Moderate-engagement (b)</td>
<td>20</td>
<td>75.95</td>
<td>8.91</td>
<td></td>
<td>b &gt; c</td>
</tr>
<tr>
<td>Low-engagement (c)</td>
<td>10</td>
<td>65.10</td>
<td>6.88</td>
<td></td>
<td>a &gt; c</td>
</tr>
</tbody>
</table>

Note. *p < .001

Further, iSAT analysis highlighted how a group of similar achievers in the pre-test performed in the post-test. Figure 3 provides the overall transition pattern among different levels of learners. There were 7 participants (17.5% of total) who improved from the low in the pre-test to the medium in the post-test. There were 6 participants (15% of total) who improved from the medium in the pre-test to the high cohorts, respectively, in the post-test.

![Figure 3. Stratified Attribute Tracking diagram for the analysis of different levels of learning achievers.](image-url)
The overall distribution data of learning time, learning achievement, and learning engagement to view the general trends in Figure 4.

![Figure 4](image)

**Figure 4.** The distribution data of learning time, learning achievement, and learning engagement.

4.2 Analysis of learning engagement of the LA e-learning

In this study, a correlational analysis not only concerning reading time and performance but also with the number of events (learning behavior) and performances. Table 3 shows the correlations between learning behaviors and learning achievement. Learning achievement was significantly correlated with reflection; reflection was significantly correlated with question replying and reading; reading was significantly correlated with scanning and question replying; question replying was significantly correlated with scanning, and skimming was significantly correlated with scanning.

<table>
<thead>
<tr>
<th></th>
<th>Scan</th>
<th>Skim</th>
<th>Question Reply</th>
<th>Reading</th>
<th>Reflection</th>
<th>Achievement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scan</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skim</td>
<td>0.519**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question Reply</td>
<td>0.318*</td>
<td>0.141</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td>0.309*</td>
<td>0.212</td>
<td>0.908**</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reflection</td>
<td>0.242</td>
<td>0.275</td>
<td>0.589**</td>
<td>0.615**</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Achievement</td>
<td>0.057</td>
<td>0.089</td>
<td>0.195</td>
<td>0.179</td>
<td>0.328*</td>
<td>1</td>
</tr>
</tbody>
</table>

Note. *p < .05, **p < .01.

5. Discussion and conclusions

In this study, the LA e-learning approach was developed, and the experiment was employed to explore the learning impact of the LA e-learning approach in a first grade English class at a senior high school in Japan. The effects of different levels of engagements were first investigated. Not only a correlational analysis for reading time and performance but also with the number of events (learning behavior) and performance were conducted. The iSAT analysis specified that more participants improved from low to medium and medium to high in the post-test. This indicates that the LA e-learning approach can guide low and medium learners to increase their learning achievements. These results suggest that the LA e-learning approach does help with learning achievement gain. The LA e-learning approach was most effective in terms of improving learners’ achievements, depending on the participants’ levels of proficiency. Consequently, it is necessary to integrate some learning strategies to optimize the use of technology in e-learning. Hsieh and Huang (2019) suggested that learning approaches should be cautiously designed and pedagogically applied to suit learners’ proficiency levels.

Furthermore, the experimental results showed that the high-engagement group had higher scores on the learning behaviors than the low-engagement group. Bonafini, Chae, Park, and Jablokow’s (2017) study also found that students’ engagement in the discussion board and video lecture increased the possibility of learning achievement. Moreover, Jamaludin and Osman (2014) indicated that learning engagement could support active learning and increase learning outcomes. Additionally, an analysis showed that significant correlations existed between learning behaviors and learning achievement. Consequently, it
is concluded that employing the LA e-learning approach, and students’ learning behaviors were directly related.

However, there are several limitations to the present study that should be noted. First, the e-learning system was designed to provide learning materials and collect the learning logs, so the learners’ autonomous learning is essential in using the system. The total time spent on learning in the research only includes the time using the system, so learners’ learning time without using the system was not able to assess. Second, the experiment lasted merely four weeks. The participants might not have practiced the learning analytics-based technique for long enough to acquire the strategies. Third, the proposed approach was only applied to a high school course; the findings might not be able to represent the effectiveness of the same approach for a longer period and different education levels. The current study focused only upon a class of high school students using an e-learning system in a single English course. In addition, a limited type of engagement was evaluated and discussed; other types of engagement, such as psychosocial processes involving students’ cognitive and affective dimensions and organizational culture (Kahu, 2013) were not addressed in this study. Accordingly, the researchers of this study intend to expand their future research by adding different types of engagement in the experimental study.

To sum up, the LA e-learning approach was integrated into an e-learning system to allow students to engage in autonomous learning as well as to have positive learning behaviors in the system. For future research, it would be valuable to explore the effectiveness of with or without the LA e-learning approach in higher education of language learning courses. It would also be worth measuring the impacts of the approach from various aspects, such as students’ higher-order thinking, self-efficacy, and self-directed learning. In addition, the investigation of students’ cooperative behaviors, as well as their interactive behaviors before class, in class, and out of class, could be valuable. Further research can also probe how to use different learning content to promote low-engagement students’ involvement in improving their learning. Finally, further investigations into e-learning systems in listening speaking or writing classrooms can provide diverse pedagogies of English language teaching development.

Acknowledgements

This work was partly supported by JSPS Grant-in-Aid for Scientific Research (S)16H06304, NEDO Special Innovation Program on AI and Big Data 18102059-0 and SPIRITS 2020 of Kyoto University.

References


Trends of E-Book-Based English Language Learning: A Review of Journal Publications from 2010 to 2019

Yuko TOYOKAWA, Mei-Rong Alice CHEN, Rwitajit MAJUMDAR, Gwo-Jen HWANG, & Hiroaki OGATA

Abstract: In recent years, e-books have been spotted in language learning research. A literature review is adopted to detect some gaps in specific areas in this study. Specifically, this paper reviewed the language learning with e-books studies published from 2010 to 2019 in selected journals from the Web of Science to explore the features and trends. A total of 43 selected studies were manually and systematically analyzed, revealing the following outcomes. The results indicated that 1) the most common participants are from preschools and elementary schools, but rarely from the secondary education level. 2) Preparatory e-books are the most commonly used. 3) Guided learning is most often adopted with e-books. 4) Reading is the primary target skill. 5) Nearly a half of the research results were positive, and nearly half were mixed. In order to support pedagogical practice with e-books, analyzing the effectiveness of functions or systems embedded in e-books would be necessary. More findings are provided to examine the usability and functionality of systems or functions in e-books and to assist teacher instructional guidance.

Keywords: Literature reviews, e-book, language learning, learning behavior

1. Introduction

The effectiveness of using technology in a learning environment has long received attention. In particular, owing to the advancement and popularity of mobile and wireless communication technologies, learners can access digital resources across real-world contexts without being limited by location or time (Andujar & Hussein, 2019; Daungcharone et al., 2019; Elaish et al., 2019). This implies the possibility of linking formal and informal learning as well as opportunities for fostering students to become active learners (Wang et al., 2020). In this case, the conventional teaching methods in the classroom are no longer applicable; instead, ubiquitous learning has attracted great attention, and e-books are used in the learning environment (Al-Harthi et al., 2020; Jia & Chen, 2020; Zhang et al., 2020).

Previous research has been conducted to discover trends in language learning using technology (Hwang & Fu, 2019; Hung et al., 2018; Zhang et al., 2020). Several key issues have been raised, such as research methods, research questions, languages, learner types, and learning outcomes. Hwang and Fu (2019) investigated trends in the research design and application of mobile language learning, and analyzed those issues. Hung and his colleagues (2018) conducted a review study and examined those critical aspects to analyze the trends in digital game-based language learning. Zhang and her colleagues (2020) did a systematic review of e-book-based language learning. It was found that the application of technology promotes language learning. These studies primarily focused on fostering learning; however, the issues of the analysis tool, dashboard, or learning logs embedded within the e-books were neglected. This study reviewed the e-book-based English language learning studies published from 2010 to 2019 in selected journals to pursue the trends of English language learning with technology specifically. Figure 1 depicts the relation of this review to the learning contexts of e-book-based English language learning, targeting learning skills and learning effects.
In light of the main goals, the current review study explores the following three research questions in the context of e-book-based English language learning from 2010 to 2019:
1. What were the learning context (academic level, technology support, and pedagogy) of the e-book-based English language learning studies?
2. What were the targeted skills of the e-book-based English language learning studies?
3. What were the effects on learning indicators of the e-book-based English language learning studies?

2. Method

2.1 Search terms and articles

First of all, keywords were identified based on the keywords from the previous review studies to search for potential articles or reviews. The searching keywords for e-books are based on Gupta and Dhawan (2019): “ebook” or “e-book” or “electronic book” or “online book” or “digital book.” The keywords for English language learning are: “language learning” or “literacy learning” or “second language learning” or “foreign language learning” or “L2” or “reading” or “listening” or “speaking” or “vocabulary” or “writing” or “grammar” or “pronunciation” (Hwang & Fu, 2019; Fang et al., 2019). In addition, “learning behavior” or “log” were added for further understanding of how previous studies collected and analyzed data.

The timespan for the most recent 10 years from 2010 to 2019 for this study was based on Hwang and Tsai’s (2011) review study and their inclusion and exclusion criteria in order to effectively investigate the trends of educational technology. Document type was decided on categories of “article” or “review.” Moreover, articles or reviews were selected from the Social Science Citation Index (SSCI) in the Web of Science (WOS) since those are considered as reliable and highly qualified, and provide stringent criteria similar to the relevant reviews for this study. There were 95 publications from 67 journals that met the criteria. It was found that research on e-books has been continuously implemented over the decade, with increases in the number of studies standing out in 2012, 2017, and 2019. For the year of 2012, six publications out of 17 were regarding e-book usage relevant to eye movement or related to physical features, while five were about library issues. In terms of the year 2017, seven out of 15 studies focused on e-book accessibility or behavior among young children, while five put the focus on media, marketing, or consumer issues. As for the year 2019, there were more issues about reading behavior or engagement (n=6) and literacy (n=3).

The studies on observing general perceptions, such as attitude toward e-books, consumer perceptions, and other matters, were excluded since the focus was on e-book-based English language learning. Furthermore, two experienced linguists carefully examined all journal articles and their abstracts. Finally, there were 43 selected for this study. The number of e-book-based English language learning studies from 2010 to 2019 is shown in Figure 2.
It can be seen that the number of publications significantly increased from the first 5 years (i.e., 15 publications) to the second 5 years (i.e., 28). This implies that e-books have been widely adopted for language learning owing to the popularity of mobile and wireless communication technologies, as indicated by several researchers (Andujar & Hussein, 2019; Daungcharone et al., 2019; Elaish et al., 2019).

2.2 Coding Scheme

Two researchers and a senior English teacher worked on the coding process to understand the trends of e-book-based English language learning settings for the last 10 years. They determined the coding scheme for this study to resolve the concern of coding reliability. The inter-rater reliability of researchers’ coding was 72%. That included learning context, targeted skills, and learning indicators. The coding scheme was modified to executing reviews as designed by Arslan (2020), Fang et al. (2019), Hung et al. (2018), and Hwang and Fu (2019). The main theoretical basis of the coding scheme was tailored from Hwang and Fu (2019). A coding scheme for e-book-based English language learning consisted of three coding items: academic level, technology, and pedagogy. Each category with its subcategories is outlined in Table 1.

Table 1. A Coding Scheme for E-book-based English Language Learning

<table>
<thead>
<tr>
<th>Contents items</th>
<th>1. Learning context</th>
<th>1.1 Academic level</th>
<th>preschool</th>
<th>elementary</th>
<th>secondary</th>
<th>higher</th>
<th>teacher</th>
<th>mixed</th>
<th>unspecified</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.2 Technology support</td>
<td>preparatory</td>
<td>researcher-developed</td>
<td>unspecified</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.3 Pedagogy</td>
<td>guided learning</td>
<td>inquiry learning</td>
<td>collaborative</td>
<td>concept/mind map</td>
<td>peer evaluation</td>
<td>peer tutoring</td>
<td>project-based learning</td>
<td>game-based learning</td>
</tr>
<tr>
<td>2. Targeted skills</td>
<td>listening speaking</td>
<td>vocabulary</td>
<td>grammar</td>
<td>pronunciation</td>
<td>integrated skills</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Learning indicators</td>
<td>positive negative</td>
<td>mixed</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2.2.1 Learning context: Academic level, Technology support, and Pedagogy

The academic levels include preschool, elementary, secondary/junior and senior high school, higher education, teachers’ education, mixed, and unspecified. Two or more learners’ academic levels mentioned in the studies were coded as mixed. The technology support is also in the learning context category. It includes preparatory e-books, researcher-developed e-books, and unspecified. The third category is pedagogy, which was divided into guided learning, inquiry learning, collaborative, concept/mind map, peer evaluation, peer tutoring, project-based learning, game-based learning, self-directed learning, mixed, and unspecified.

2.2.2 Targeted skills

This category was classified into eight coding items, namely listening, speaking, reading, writing, vocabulary, grammar, pronunciation, and integrated skills.

2.2.3 Learning indicators

This category was coded by using four coding items: positive, negative, mixed, and not available. The term “positive” indicated preferences and pleasant consequences of e-book learning, and rewarding and desirable outcomes such as resulting in effective learning and analysis. Negative indicated adverse outcomes such as non-preference and outcomes which did not result in effective learning. The term “mixed” indicated that there were both positive and negative results. The term “not available” referred to neither positive nor negative.

3. Results

3.1.1 Academic levels

Regarding the learners’ academic levels, over 20% of the research was conducted observing elementary education (n=9) followed by preschool (n=8). E-book learning for higher education also attracted researchers’ attention (n=6). A few studies were conducted for secondary education (n=2) and teachers (n=1). “Mixed” (n=7) included studies with preschoolers and parents’ combination (n=5), elementary students and teacher combination (n=1), and both elementary and secondary education (n=1). Students (n=8), children (n=1) and participants (n=1) whose ages or educational levels were not identified were coded as “unspecified” (n=10). By combining the results of preschool education and those of preschoolers and adults’ combination from “mixed,” 13 studies were on early childhood language learning with e-books.

![Figure 3](image)

*Figure 3. Academic levels of e-book-based English language learning studies from 2010 to 2019.*

The result shows that language learning with e-books is popular for early childhood education, while it is not so popular at the secondary school level. In addition, how e-books were used was
observed: used for independent reading, shared-reading, or adult-led reading. There were 11 cases out of 13 which found that they were used for shared reading (nearly 85%). Some studies examined multiple cases, such as both shared-reading and independent reading. Therefore, the number did not match the total number of 13; rather, it exceeded 13. Three independent reading and three adult-led reading cases were found. The results indicate that e-books were used for learning and playing with, and were used as communication tools for young children. Parents and teachers promoted the development of the reading process during shared reading.

As the academic level increases, e-books were used more as textbooks. Some researchers developed or used e-books with functions such as computer games, concept mapping systems, and reading guidance and annotation maps to support learning (Smith et al., 2013; Li, 2015; Li et al., 2014). There was a study that observed teachers’ pre-operational behavior toward e-book learning. Adequate instructional assistance for teachers would be necessary to support them in effectively using e-books as learning tools. As for adults, there was no attention paid to their language learning with e-books in the selected studies.

3.1.2 Technology support

E-books were examined in terms of the availability of technology support their use. The primary purpose was to find whether the e-book systems used for the studies were preparatory, researcher-developed e-books, or unspecified (Figure 4). There were 32 preparatory and seven researcher-developed e-books. Four studies did not mention the types of e-books, so they were coded as “unspecified.”

![Figure 4. Technology support of e-book-based English language learning studies from 2010 to 2019.](image)

Some functions or systems which helped to promote learning were observed among the preparatory e-book systems. According to Lin and Lin (2012), research participants found mark-up tools such as highlighting helpful, and pictures/animations and music/sound effects increased their motivation. E-book-based language learning in an informal learning environment was observed by Hwang et al. (2015) by using annotation-sharing functions. They suggested that e-book-based learning helped to reinforce learning by seamlessly connecting formal and informal learning. There were several studies conducted with analysis tools, either embedded or used separately. Some of them came with reading behavior or concentration monitoring systems and eye-tracking or reading rating systems. Hsu et al. (2012) developed a reading concentration monitoring system for use with e-books to examine students’ reading concentration rates. The functions provided researchers with the opportunity to evaluate the data during the reading process in real time. Boticki et al. (2019) used a log as a part of the learning analytics to examine user models. By using functions or systems in e-books, researchers could collect detailed data and apply the evidence to promote learning performance.

The objective of the system was also examined for the e-books developed by researchers. Two e-books were developed for enhancing learners’ motivation and engagement while improving language proficiency. For example, one study innovated a role-play picture e-book to boost learners’ motivation and reading engagement by using the effect of emotive selfies (Kao et al., 2019). Two e-books were designed to analyze the reading process or behaviors (Lin et al., 2019; Li et al., 2014). In terms of the analysis tools, there were two e-books with the dashboard to monitor and analyze learning performance, and one used a log as a part of the learning analytics (Wu, 2016). Examining the usability, functionality, and effectiveness of the systems in e-books, as well as understanding learning processes to optimize learners’ learning by using e-books should be taken into account more often in the future.
3.1.3 Pedagogy

Regarding pedagogy, guided learning was adopted the most with the use of e-books (n=14). Two types of guided learning were observed; one was human-guided learning (n=8), and the other was e-book guided learning (n=6). Five studies adopting self-interaction or individual readings were coded as self-directed learning. Three studies adopted project-based learning, one adopted concept mapping, and one adopted game based-learning as the learning strategies. Eight studies in which multiple learning strategies were adopted were coded as “mixed”: six guided-learning and self-directed learning integration, and two guided learning and collaborative learning integration. The 11 studies which did not mention any learning methods were coded as “unspecified” (Figure 5).

It was found that e-books were used in the same way as conventional reading materials for language learning, and using e-books for shared reading guided and supported by teachers or parents has been popular, especially for early literacy.

Collaborative learning itself and peer evaluation or tutoring were not observed in this study. Those learning strategies would be more likely to be adopted when using a system that utilizes logs from which learners’ information can be extracted to form groups or peers who share similar interests or academic levels. Elaborately designed teaching activities or learning strategies using e-books with such tools or functions would open up diverse learning opportunities in the near future.

3.2 Targeted skills

Regarding target language skills, over 76% of the studies (n=33) were conducted on reading skills. There was a single study on vocabulary learning (n=1), while nine studies were conducted to examine integrated skills; those were the integration of listening and vocabulary, reading and vocabulary, speaking, reading, and vocabulary, listening and reading, and writing and oral skills. Preschoolers’ alphabet and letter recognition were categorized as reading, and phonological awareness was categorized as listening. Recall and cognitive outcomes were categorized as reading, spelling as writing, talk and vocalization as speaking, and letter-related behaviors such as saying names of letters or objects were coded as vocabulary.

Regarding the reading skills, 29 out of 33 studies examined reading behavior/attitude, engagement, or preference; data for those elements were collected mostly by implementing observation (including video recording), questionnaires, or quizzes. Thirteen studies analyzed reading performance, such as reading comprehension, fluency, recognition, or reading rates, and the data for those elements were collected mostly by using tools or functions embedded in the e-books. Some studies used both (or several) data collecting methods.

For integrated skills, e-book reading enabled users to learn not only reading but also other language skills like listening, speaking, writing, and vocabulary synchronically. One study (Smith et al., 2013) was on vocabulary skills. The vocabulary was integrated with other elements such as listening and reading in the study. Moreover, there were no studies targeting grammar skills found in this review.
3.3 Learning indicators

In terms of learning indicators of the studies’ results, nearly half of the research was found to be positive (n=20, 46%), and the other half was mixed (n=20, 46%), which included both positive and negative results. Only two research results were found to be negative (n=2), both of which were conducted to compare the differences between e-books and p-books as learning tools. Their results indicated that students preferred p-books to learn rather than e-books. Students reported that they could attain better focus and retention of information in printed formats (Mizrachi et al., 2018), even though they were satisfied with e-books. P-books also appeared to enable better reading comprehension (Jeong, 2012).

Positive results showed that e-books were not just electronic reading tools; rather, they were well designed and elaborated textbooks with advanced technologies that enabled learners to increase their interest in reading, and improve their language skills. Not only integrating elaborated functions or systems into e-books and implementing those, but also providing instructional guidance for teachers to support their pedagogical skills with e-books will be essential to make e-book learning optimal. Overall, learners are satisfied with the use of e-books. Motivation and interaction are the most critical factors that support e-book language learning.

4. Conclusions and discussion

Major findings for the recent 10-year review studies on e-book-based English language learning in the selected journals from SSCI are summarized as follows. The learning context of e-book-based English language learning studies indicated that studies on e-books for preschool or early childhood education appeared to gain attention. There were few studies conducted at the secondary school level. As for the technology support of the e-books, preparatory e-books were the most common learning tools in the studies, while some researcher-developed e-books were implemented. Guided learning was adopted the most when using e-books. The majority of the recent studies were implemented on examining reading skills, while some others analyzed integrated skills. The targeted skills of e-book-based English language learning studies from 2010 to 2019 found that reading comprehension was examined throughout the decade; while early studies observed reading rate, more of the recent studies observed measuring the early literacy process. The effects on the learning indicators of e-book-based English language learning studies revealed that nearly half of the studies concluded with positive and nearly half with mixed results.

Several limitations of the study need to be acknowledged. First, research methods should have been more closely observed since research design is one of the crucial issues for the studies. Furthermore, the trends of e-books for language learning on the use of technology were not sufficiently discussed. Further attention to specifically designing and developing e-books to collect and analyze data would help to visualize the effectiveness of using e-books for learning and overall reformation of the educational system. By using functions or systems in e-books, researchers can collect detailed data and prove the effectiveness of using e-books for learning as evidence. Moreover, the use of e-books would require more guidance and teaching guidance to support teaching practice.

Acknowledgements

This work was partly supported by JSPS Grant-in-Aid for Scientific Research (B)20H01722, JSPS Grant-in-Aid for Scientific Research (S)16H06304 and NEDO Special Innovation Program on AI and Big Data 18102059-0, SPIRITS 2020 of Kyoto University, and the Ministry of Science and Technology of the Republic of China under contract numbers MOST 106-2511-S-011 -005 -MY3.

References


**Coded papers**


Using Computer-like Rules to Give Automatic Grammatical Written Corrective Feedback

A Case Study of Structural Particles

“的”, “地” and “得”

Si Hui LIU*, Ming Ming CHIU*, Zhan WANG* & Sze Ming LAM*
*Assessment Research Center, the Education University of Hong Kong, Hong Kong
* victoria1155081866@gmail.com

Abstract: Mandarin speakers often confuse three structural particles “的”, “地” and “得” in their formal writings and Chinese teachers often correct these errors in students’ essays. To reduce Chinese teachers’ laborious workload, the present research aims to investigate how to use technology to efficiently correct grammar errors with feedback. All three structural particle instances were analyzed in 50 primary one level HK students’ essays. Their grammaticality judgments were based on three rules derived from literature: (1) A+的+B. B is always a Noun. A is used to modify B; (2) A+地+B. B is always a Verb. A is used to modify B to indicate how; (3) A+得+B. A is always a Verb. B is used to modify A to indicate how. 700 out of 804 instances of “的”, “地” and “得” were identified as structural particles. By using automatic Chinese parsing software NLPIR-ICTCLAS (see Zhang, 2018), 80% could be judged easily and accurately while the rest required a deeper analysis, which indicates that automatic Part of Speech (PoS) judgment is quite useful and helpful to Chinese teachers’ grammatical feedback on these three structural particles in students’ essays. Difficulties and problems in grammaticality judgment from a deep level were further analyzed. Corresponding solutions and directions for future research were also discussed. This research can shed light on how to use Part of Speech (PoS) parsing software to facilitate automatic grammatical written corrective feedback and how technology can enhance language learning and teaching in a Chinese language education context.

Keywords: grammar, automatic, Written Corrective Feedback (WCF), Chinese, structural particles, 的, 地, 得

1. Introduction

“的”, “地” and “得” are three structural particles of high frequency due to their active role in sentence structure formulation. However, they are often easily mixed up by Mandarin speakers in their formal writings. In daily conversations, Mandarin speakers do not need to distinguish one with another since they share similar pronunciation “-de” in Mandarin speaking. However, in formal writing, these three words are used in different ways, especially when they collocate with other specific words in the discourse. The confused uses of them in Chinese formal writings decrease both language expression accuracy and people’s communication efficiency.

Due to the above situation, Chinese teachers highly emphasize the distinctions among these three structural particles in their daily teaching, especially through giving direct grammatical corrective feedback on students’ essays. However, such grammatical written corrective feedback can be ineffective or inefficient. From students’ perspective, on the one hand, with only pure error corrections but insufficient corresponding grammar knowledge scaffolding, students still do not know how to use these three structural particles accurately, and thus might make these errors again in their future writings. On the other hand, learners need to go through the language learning process from knowing a new
grammatical knowledge to using it in their natural communication accurately, since it requires practicing efforts to crystallize such new knowledge and to integrate it with their prior knowledge. Thus, in some cases, learners keep making grammatical errors until they reach the final crystallization stage.

As a result, Chinese teachers have to correct such errors in their students' essays again and again so as to help their students with the grammatical knowledge gap identification and new grammar knowledge crystallization, which increases their working burden due to more grammatical written corrective feedback. To reduce Chinese teachers’ laborious workload, the present research aims to investigate how to use technology to give grammatical written corrective feedback in an efficient and economic way.

We suggest the following procedure: first, identify these three structural particles “的”, “地” and “得” in students’ essays; next, find corresponding collocative words occurring before and after these three structural particles “的”, “地” and “得” within each discourse instance; then, apply the following three rules derived from previous relevant literature (A represents collocative words occurring before these three structural particles within the discourse whereas B represents collocative words occurring after them) centering on Part of Speech (PoS) with the support of automatic Chinese parsing software NLPIR-ICTCLAS (see Zhang, 2018):

1. A+的+B. B is always a Noun. A is used to modify B;
2. A+地+B. B is always a Verb. A is used to modify B to indicate how;
3. A+得+B. A is always a Verb. B is used to modify A to indicate how.

The present research might shed some light on how to use technology to facilitate grammatical written corrective feedback in Chinese learning and teaching and beyond.

2. Literature Review

2.1 “的”, “地” and “得” as Structural Particles in Formal Writing

According to the literature review, previous research on “的”, “地” and “得” as structural particles in Chinese formal writing mainly centers on answering two questions. One, how “的”, “地” and “得” should be used grammatically as structural particles in formal writing. Two, how “的”, “地” and “得” are actually used in different kinds of language situation contexts.

2.1.1 How They Should be Used Grammatically as Structural Particles in Formal Writing

Studies focusing on how “的”, “地” and “得” should be used grammatically as structural particles in formal writing, mainly discussed their standard usages centering on their usage similarity and difference (e.g., Peng & Liu, 2013; Wang, 2010). A large group of literature was written to facilitate Chinese language teaching and learning within the education context. The audience for these studies mainly are primary school students (e.g., Chen, 2011; Ding, 1996; Fu, 2014), secondary school students (e.g., Huo & Zhang, 2015; Wang, 2010), and their teachers. These studies pointed out the high frequency and high probability of confusion by students when using these particles while learning Chinese.

Thus, the above mentioned studies present very clear and fine-grained grammatical explanations on how to use “的”, “地” and “得” as structural particles accurately in formal writing, aiming at facilitating students’ better understanding and command of their grammatical usages.

2.1.2 How They Are Actually Used in Different Kinds of Language Situation Contexts

The second group of studies focuses on how “的”, “地” and “得” are actually used in different kinds of language situation contexts. These studies usually center on the commonly mixed usages of these structural particles and summarize the common error patterns in contrast with the standard usages (e.g., Du, 2011; Liu, 2006; Luan, 2020; Zuo, 2008). Studies were also conducted in different specific discourse context such as in medical discourse (e.g., Xu, 1988) and editing discourse (e.g., Yang, 2017),
which contribute to a more comprehensive picture of all error patterns, especially within these special languages using contexts.

Studies derived from Teaching Chinese as a Second Language is one important component of this group of literature. Usually, scholars summarize the erroneous usage patterns made by Chinese L2 learners and analyze these errors for their error type categorization and further analyze why L2 learners made these errors by using relevant Applied Linguistics Theories such as Second Language Acquisition, Inter-language, Error Analysis and so on (e.g., Deng, 2011; Guo, 2011; Wang, 2020). Useful and practical learning and teaching advice are derived from these research to facilitate Teaching Chinese as a Second Language on issues like textbook and follow-up exercise compilation (e.g., Ma, 2013), curriculum design (e.g., Zhou, 2011), teaching strategy improvement (e.g., Yang, 2014) and so on.

2.2 Written Corrective Feedback (WCF) in Chinese Writing

Compared with other language (such as English) contexts, studies on Written Corrective Feedback (WCF) on Chinese writings are very small in number. Fang and Wang (2019) investigated teachers’ feedback to Chinese L2 learners’ writings and students’ corresponding responses in their writing revisions. Within a computer-mediated WCF context, Hsieh, Hiew and Tay (2017) studied how Chinese teachers adopted an automated essay marking system to give instant WCF towards students’ writings. Han (2019) investigated how teachers used an online system to offer WCF. Both studies explored Chinese L2 learners’ perceptions and views towards teachers’ WCF on their writings. Centering on facilitating self-identification and self-correction of grammar errors, Ai (2017) conducted research tracing how feedback dynamically changed during the corrective feedback process on the ba-construction (“把”) in Chinese L2 acquisition within the context of computer-assisted language learning.

2.3 Automatic Chinese Grammatical Written Corrective Feedback (WCF)

Many studies use computational models, statistical analyses or machine learning to detect Chinese grammar errors, correct them or give corrective feedback (e.g., Chang, et al., 2014; Lee, Tseng, & Chang, 2019; Ren, Yang, & Xun, 2018; Yeh, et al., 2015). These studies require extensive technical expertise, and hence are not comprehensible to most Chinese teachers. Hence, this line of research has a gap: few studies that investigate how technology can help primary or secondary school Chinese teachers with little technical expertise teach their students to detect errors, correct them, understand the correct and incorrect grammatical uses, and reduce such errors in their future writing. Thus, the present research aims to fill this niche by taking a step towards Part of Speech (PoS) parsing computer software to facilitate Chinese teachers’ grammatical Written Corrective Feedback (WCF) on the use of three structural particles “的”, “地” and “得” in students’ essays.

3. Research questions:

For Chinese teachers’ grammatical written corrective feedback on the three structural particles “的”, “地” and “得” in students’ essays:

1. To what degree can automatic part of speech (PoS) parsing computer software help with grammaticality judgments?
2. What kinds of difficulties (problems) occur during the process mentioned in 1?
3. What advantages and disadvantages of automatic part of speech (PoS) parsing compared with traditional written corrective feedback (WCF)?

4. Methodology

4.1 Data Collection
After getting the ethics approval from the local schools, we collected Chinese essays written by primary one level students from HK local schools and randomly selected 50 essays for our research data. All collected essays were scanned in PDF format first and transcribed into the TXT file later. The 50 essays were automatically segmented and parsed by using the NLPIR-ICTCLAS software (Zhang, 2018) and later input into one Microsoft Word File.

First, we searched for Chinese characters “的”，“地” and “得” in the Microsoft Word File and highlighted all of them. Next, we read through the discourse and judged whether “的”，“地” and “得” used as a structural particle (e.g., rather than a part of a content word) within each collected instance. Then, we collected all structural particles using instances (each complete sentence) and input them into three spreadsheets in Microsoft Excel respectively. Considering the grammaticality judgment in this research was PoS-sensitive, we separated “的”，“地” and “得” from the occurring collocative words before and after these three structural particles within each instance like this in the Excel spreadsheets respectively:

Table 1. An Example of Our Data Collection by Using the NLPIR-ICTCLAS Software (Zhang, 2018)

<table>
<thead>
<tr>
<th>小明/nr2</th>
<th>的/ude1</th>
<th>父母/n</th>
</tr>
</thead>
<tbody>
<tr>
<td>柏/rr</td>
<td>柏/ude1</td>
<td>柏/n</td>
</tr>
<tr>
<td>她/tr</td>
<td>的/ude1</td>
<td>名字/n</td>
</tr>
<tr>
<td>她/tr</td>
<td>的/ude1</td>
<td>名字/n</td>
</tr>
</tbody>
</table>

4.2 Data Analysis

According to the three rules derived from previous relevant literature (A represents collocative words occurring before these three structural particles within the discourse whereas B represents collocative words occurring after them): 1). A+的+B. B is always a Noun. A is used to modify B; 2). A+地+B. B is always a Verb. A is used to modify B to indicate how; 3). A+得+B. A is always a Verb. B is used to modify A to indicate how.

We introduce the NLPIR-ICTCLAS software (Zhang, 2018) to help with parsing our essay corpus data as follows:

Table 2. An Example of Our Data Analysis by Using the NLPIR-ICTCLAS Software (Zhang, 2018)

<table>
<thead>
<tr>
<th>善良/a</th>
<th>的/ude1</th>
<th>心/n</th>
<th>(kind heart)</th>
</tr>
</thead>
<tbody>
<tr>
<td>快快/d</td>
<td>吃/v</td>
<td>葉/ule</td>
<td>(fastly ate)</td>
</tr>
<tr>
<td>模仿/v</td>
<td>得/ude3</td>
<td>最/d</td>
<td>相似/a</td>
</tr>
</tbody>
</table>

Each alphabetic symbol following the “/” indicates the Part of Speech (PoS) automatically parsed by the computer software: “a” indicates adjective, “n” indicates noun, “d” indicates adverb, “v” indicates verb and “ule” indicates past tense. “的”，“地” and “得” are parsed as ude1, ude2 and ude3 respectively due to their different grammatical usages and functions.

We searched for the following 3 patterns in three spreadsheets respectively: 1). 的 + Noun; 2). 地 + Verb; 3). 遠 + 得. Next, we separated them into those that followed this pattern versus those that did not. For the latter, we first judged whether they were grammatically right and separated the right from the wrong ones. Then, we explored the relevance of the patterns of each instance and put instances with similar patterns into the same group. Finally, we described the sentential pattern for instances within each group and conducted statistical analysis on frequency and percentages counting by using relevant functions in Microsoft Excel.

5. Results and Discussions

5.1 Focusing on “的”
Among the 646 collected “的” usages, 591 were identified as structural particles. 471 instances satisfied the rule “A+的+B. B is always a Noun. A is used to modify B.”, which had the highest frequency. One instance was identified as grammatically wrong since B is a verb within this instance whereas it should be a Noun according to the rule. For the other 119 instances, all of them are grammatically right but only satisfy the rule for a deeper further analysis of the sentence structure beyond the surface level. There are two reasons for this rule satisfaction inconsistency.

Firstly, the PoS problem, namely, computers have problems in judging the PoS of B if there are any modifiers before B with other PoS or B is a multi-category word. For the above mentioned 471 instances satisfying the rule, B is always one specific pure noun. However, there are also other circumstances. One situation is an inserted modifier between “的” and B, (69 of 79 instances for this PoS problem). In such a situation, the inserted modifier and B work together to form a whole phrase. B determined the PoS of the whole phrase as the head of the phrase. Thus, the PoS of the whole structure after “的” was still a noun, which allowed it to still satisfy the rule. For example, within the structure “我/rr 的/ude1 好/a 朋友/n (my good friends)”, adjective “好/a (good)” is used to modify noun “朋友/n (friends)”. The PoS of the head “朋友/n (friends)” determined the PoS of the whole phrase “好/a 朋友/n (good friends)” as noun (rather than adjective). Thus, “我/rr 的/ude1 好/a 朋友/n (my good friends)” satisfies the rule “A+的+B. B is always a Noun. A is used to modify B.” Another situation is the multi-category word, which accounted for 10 out of the 79 instances of the PoS problem in total. Since such words had more than one PoS categories, the PoS might be judged as another category rather than the one that fitted into the context, which imposed a negative impact on rule judgment. For instance, in “春/tg 風/n 的/ude1 陪伴/vn (the companionship of spring breeze)”, the multi-category word “陪伴/vn (companionship)” might be parsed as a verb (but is actually used as a noun in this context), which might mislead people to judge it as breaking the rule “B is always a Noun.”

Secondly, the multiple rules application problem: when multiple grammar rules are applicable but incompatible, computers have difficulty choosing which rule to apply. These data had 40 such instances, mostly in the following two situations at the discourse level. One situation was the judgment sentential pattern “......是/會......的” (...be-Verb/will...), 27 of 40 instances. For instance, “媽/n 媽/n 總/n 是/vshi 和/cc 蕭/n 可/v 親/n 的/ude1 (Mom is always amiable)” did not satisfy the rule “A+的+B. B is always a Noun. A is used to modify B.” as there was no such a B in this sentence. However, it
satisfied the judgment sentential pattern “......是/會......的” (... be-Verb/will...) and thus was still grammatically right. The other situation was omission at the discourse level, 13 of 40 instances. For instance, in the sentence “而/cc 牠/n 最/d 愛/n 吃/v 的/ude1 是/vshi 野果/n 和/cc 種/n 果/ndg. (And its favorite food are wild fruits and seeds.)”, the noun between “的” and “是” (be-Verb) was omitted, so this sentence did not satisfy the rule “A+的+B. B is always a Noun. A is used to modify B to indicate how”. However, if we judge from a further deeper discourse/sentential (rather the surface level), all instances included in the multiple rules application category were grammatically right, which also seemed indicate that when multiple rules applied for the same sentences, the deeper rules (e.g., discourse/sentential rules) tend to be prioritized over surface rules (e.g., lexical collocation rules).

5.2 **Focusing on “地”**

### Table 4. Results for Structural Particle “地”

<table>
<thead>
<tr>
<th>地</th>
<th>69</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satisfy the rule &quot;A+地+B. B is always a Verb. A is used to modify B to indicate how&quot;.</td>
<td>59</td>
</tr>
<tr>
<td>e.g., 快快/d 地/ude2 吃/v 了/u 早餐/n (ate breakfast quickly)</td>
<td></td>
</tr>
<tr>
<td>Reasons identified for not satisfying the rule on the surface level (but are actually grammatically right):</td>
<td></td>
</tr>
<tr>
<td>PoS problem (PoS should be judged by the head of the structure)</td>
<td>10</td>
</tr>
<tr>
<td>1. The PoS of B should be judged based on the head of B.</td>
<td>3</td>
</tr>
<tr>
<td>e.g., 高/a 興/n 地/ude2 大/a 叫/vi (shout aloud happily)</td>
<td></td>
</tr>
<tr>
<td>2. The PoS of B should be judged based on the head of B.</td>
<td>7</td>
</tr>
<tr>
<td>e.g., 緊/n 張/n 地/ude2 向/p 她/rr 道歉/vi (apologize to her nervously)</td>
<td></td>
</tr>
</tbody>
</table>

98 instances of “地” were collected from our corpus, among which 69 used as structural particles. 59 out of 69 instances satisfied the rule “A+地+B. B is always a Verb. A is used to modify B to indicate how”. Very similar to the PoS problem with “的”, inserted modifiers between “地” and B were also identified in examples like “高/a 興/n 地/ude2 大/a 叫/vi (shout aloud happily)”. The adjective “大/a” (aloud) was used as an insertion between “地” and verb “叫/vi (shout)”, which were 3 instances in total. Another case was a complement insertion between “地” and B. For instance, in “緊/n 張/n 地/ude2 向/p 她/rr 道歉/vi (apologize to her nervously)”, “向/p 她/rr (to her)” was the inserted complement that modified the following verb “道歉/vi (apologize)” (7 instances). In both cases, the PoS problem could be solved via judging the PoS of the head of the structure after “地” in “A+地+B”.

### Table 5. Results for Structural Particle “得”

<table>
<thead>
<tr>
<th>得</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satisfy the rule “A+得+B. B is always a Verb. B is used to modify A to indicate how”.</td>
<td></td>
</tr>
<tr>
<td>1. B is always an Adj./Adv.</td>
<td>27</td>
</tr>
<tr>
<td>e.g., 看/v 得/ude3 入神/a (watch entrenched) or 活/v 潑/n 得/ude3 很/d ! (very outgoing)</td>
<td></td>
</tr>
<tr>
<td>2. B is usually a Verb. B often modifies A to indicate how--the extent of A.</td>
<td></td>
</tr>
<tr>
<td>e.g., 高/a 興/n 得/ude3 跳/vi 了/u 來/n. (jumped with joy)</td>
<td>4</td>
</tr>
<tr>
<td>3. B is a consequential clause indicating results/influence towards the agent. Could</td>
<td>9</td>
</tr>
</tbody>
</table>
not define the PoS of B as it is a clause.

For “得”, 40 structural particle uses were identified out of from the 60 collected instances. All of them satisfy the rule “A+得+B. A is always a Verb. B is used to modify A to indicate how.” and two minor groups were categorized based on the PoS of B. Within the first group, the PoS of B were all adjectives or adverbs like “看/v 得/ude3 人/a 神/a (watch entrance)" or “活/v 演/n 得/ude3 很/d ! (very outgoing)”. For the other group, B was a complement of A. The PoS of B could be a pure verb like in “高/a 興/n 得/ude3 跳/vi 了/re/vi 起/vf 来/n。 (jumped with joy)”. Sometimes, B could also be a clause or sentence like in “表演/v 得/ude3 令/n 觀/n 眾/n 樂/n 不可/v 支/q (the performance made the audience overjoyed)”. However, the latter within the second group would be very difficult to be automatically analyzed by the computers at the discourse level.

6. General Discussion

As we have discussed above, among the 804 instances of “的”, “地” and “得” extracted from 50 primary one local students’ essays, 700 were identified as structural particles. 79.86% (472 out of 591) could be successfully judged on the surface level of the function use of “的“, 85.51% (59 out of 69) for that of “地” and 77.5% (31 out of 40) for that of “得”. Generally speaking, 80.29% (562 out of 700) could be judged accurately on the surface level through using the automatically parsed PoS and the rest 19.71% (138 out of 700) needed deeper analysis, which indicates that for Chinese teachers’ grammatical feedback on the three structural particles “的”, “地” and “得” towards students’ essays, automatic PoS judgment is quite useful and helpful. The error rate is very low (only 1 out of 804, approximate 0.12%), which is inconsistent with the conclusion derived from a large amount of previous studies conducted by Mandarin speaker scholars before (e.g., Du, 2011; Fu, 2014; Gong, 2016; Huo & Zhang, 2015). We hypothesized that it is because the essays collected in this study were produced by HK primary students whose most frequently used language is their mother tongue Cantonese. Whereas these three structural particles have functionally equivalent words in Cantonese (“得” for “的”, “得” for “地” and “到” for “得”) and they do not share similar pronunciation (whereas they share similar pronunciation “-de” in speaking Mandarin). Thus, the possibility of confusing them due to their pronunciation similarity is much lower. Future research could provide further investigation on this issue with a larger data set.

Table 6. A Summary of Results for Structural Particles Analysis (Surface & Deep)

<table>
<thead>
<tr>
<th>Structural particles</th>
<th>Surface level</th>
<th>Deeper analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>591</td>
<td>472 (1 error) (79.86%)</td>
<td>119 (20.14%)</td>
</tr>
<tr>
<td>69</td>
<td>59 (85.51%)</td>
<td>10 (14.49%)</td>
</tr>
<tr>
<td>40</td>
<td>31 (77.5%)</td>
<td>9 (22.5%)</td>
</tr>
<tr>
<td>700 (100%)</td>
<td>562 (80.29%)</td>
<td>138 (19.71%)</td>
</tr>
</tbody>
</table>

Following are the two main categories of difficulties/problems that were confronted with during the automatic grammatical written corrective feedback correction process in this study:

6.1 PoS Identification

As we have discussed above, automatic PoS identification is good for its efficiency and convenience dealing with large amounts of data in automatic PoS segmentation and parsing. We identified the following specific aspects which can be further developed:

Firstly, PoS of 4 character idioms. 4 character idioms is one quite unique and special expression in Chinese language. They are not regarded as an integrated unit but as four independent Chinese
characters in automatic segmentation and parsing, which requires more manual work on PoS judging on them for the present stage.

Secondly, PoS of multi-category words. Multi-category words are words that have several different PoS. Unlike Germanic languages (like English, German etc.) which are rich in inflected forms and morphological changes, Chinese words are relatively stable in form (inflection) and the PoS is relatively reflected more in other aspects like its collocation with other words, functions in sentential patterns and so on. Thus, it is difficult for computers to parse the PoS of multi-category words accurately since they cannot judge the PoS according to the specific language usages and context. Furthermore, the PoS of the same Chinese word would even change flexibility for different specific usages which is derived from classical Chinese.

Thirdly, the PoS of phrases. Most PoS are word-based, namely, they segment the independent words based on frequency and parse the PoS accordingly. But sometimes the judged target might be beyond the word, like phrases or collocations. In such a situation, the PoS of the whole phrase or collocation should be judged based on the head of the structure. However, the machine could not realize this very well for the present as it requires semantic and syntactic analysis at a higher level.

Fourthly, the inequivalence between PoS and sentence constituents. Many grammaticality judgments are based on grammar rules, which are closely related to sentence constituents (such as subjects, predicates, objects, attributives, adverbials, complements, predicatives etc.) especially when judging at the sentential level. However, the PoS cannot correspond with them completely. For instance, the PoS of the complement could be an adjective, adverb, verb or sometimes even a clause. Thus, finding a way to bridge the gap between PoS and sentence components will be meaningful and contribute to using PoS to facilitate grammaticality judgments.

6.2 Multiple Rules Application in Grammaticality Judgment

When multiple rules could be applied to the same language phenomenon, the grammaticality judgment becomes complicated as the possibility becomes larger in number. Sometimes, we cannot judge the usage as ungrammatical when it does not fit well into one grammar rule from one language aspect (e.g., word collocation from the lexical level) since it might suit another rule from another language aspect (e.g., sentence pattern from the discourse level). Thus, how to manipulate and prioritize multiple rules within the same discourse text is a question worth investigating.

Another issue lies in the potential inconsistency between language structure on the surface level and deep level. For instance, the missing of some grammatical components on the surface level might be attributed to omission at the discourse level rather than lack of sentential components ungrammatically. For computers, it is much easier to detect the written down part on the surface level; however, it is very hard to detect and judge the omitted part, or the underlying full picture of the whole sentence structure. Thus, a more fine-grained hierarchical sentential structure analysis and semantic analysis at the discourse/sentential level is much needed for future research.

In sum, compared with traditional manual written corrective feedback given by Chinese teachers, this automatic part of speech (PoS) parsing enables written corrective feedback in an efficient way for grammaticality judgment and thus saves teachers’ time and energy by relieving them from their laborious workload. Furthermore, such methods are applicable to similar grammar errors. However, the approach requires transferring abstract grammar rules into specific ones suitable for computer processing. Further, eliminating counter-examples is critical.

7. Limitations and Future Research

This study has several limitations. Firstly, the data size is small, which might limit its representativeness. Secondly, the pronunciations of these three Chinese characters are completely different (i.e., “dik1, dei6, and dak1”) in Cantonese but are very similar in Mandarin (i.e., all three are “de”). Such pronunciation distinctions help HK students distinguish them one from another, which leads to low error rates. Future research could use larger writing data size from non-HK students (e.g., English L1 Mandarin L2 speakers).
8. Conclusion

This small-scale research focuses on the grammatical written corrective feedback on three frequently used but easily mixed-up structural particles “的”, “地” and “得” in 50 collected HK school students’ essays. An automatic Chinese parsing software NLPIR-ICTCLAS (see Zhang, 2018) was adopted to facilitate the automatic part of speech analysis for the sake of grammatical judgment of these three structural particles. Most previous research studied the written corrective feedback of these three structural particles manually, which usually summarized the error patterns and gave further elaborations with simple examples. The present research solves this problem through adopting automatic PoS computer software to facilitate grammaticality judgment. Some difficulties and problems confronted with during the automatic grammaticality judgment process are also mentioned with corresponding future direction. The present research could be regarded as an example of how to use computer software to facilitate automatic grammatical written corrective feedback towards Chinese learners’ writings, which hopefully, could shed some light on more good research for technology-enhanced language learning and beyond.

Acknowledgements

We would like to thank all the folks in the Assessment Research Center (ARC) in the Education University of Hong Kong for their warm support and gracious help all the time.

References

Han, J. (2019). The Effects and Students’ Views of Teachers’ Coded Written Corrective Feedback: A Multiple-Case Study of Online Multiple-draft Chinese Writing.


Measuring Students’ Online Language Learning Engagement: Towards the Development and Validation of a Scale

Lin LUANa,b*, Yan DONGb & Miao CAOb

a Department of Foreign Languages, Beijing University of Posts and Telecommunications, China  
b School of Educational Technology, Beijing Normal University, China  
*luanlin32@126.com

Abstract: Language educators face significant challenges of engaging language learners especially in the online course of target language learning. Low participation and disinterest in online learning activities are commonly related to low student engagement in online classes. Along with the adversity in the online course of target language learning, student engagement is considered to play a pivotal role on learners’ continuous effects and academic achievements in second language learning. Although there has been a growing interest in the conceptualization of student engagement, the literature lacks a subject-specific measuring instrument to operationalize engagement in an online setting. This study attempts to fill this gap by developing and validating an online language learning scale (OLLE) scale for university students. The sample included 454 students at a large comprehensive university in China. Both exploratory and confirmatory factor analyses were conducted to determine construct validity and reliability. Consistent with the adopted theoretical framework, the 19-item OLLE scale comprised four components: behavioral engagement, cognitive engagement, emotional engagement, and social engagement. It is also indicated that EFL learners had the highest level of behavioral engagement and the lowest level of emotional engagement in the process of online language learning. These results along with their implications within the pedagogical and research contexts are also discussed in this study.

Keywords: student engagement, scale development, online language learning, English as a foreign language (EFL), higher education

1. Introduction

Due to the COVID-19 across the world, a growing number of learning activities is taking place in online contexts such as tutoring systems or virtual learning communities. However, due to the temporal and spatial separation among teachers and learners in the technology-enhanced learning environment, students may suffer a sense of loneliness, disconnectedness and other negative feelings. Low participation (Thomas, 2002) and disinterest in online learning activities (Xie, Durrington, & Yen, 2011) are commonly related to low student engagement in online classes. Confronting the difficulties and challenges in the online course of target language learning, learners need to depend on a series of cognitive, affective, and behavioral components in order to persist and then to complete the learning tasks (Zheng, Liang, Li, & Tsai, 2018). Among these components, learners’ engagement are crucial element for enhancing learner motivation and their academic achievement (e.g. Connell, Spencer, & Aber, 1994; Fredricks et al., 2016; Wang & Holcombe, 2010).

2. Literature Review

Engagement can be viewed as active participation in the learning process, and contributes to deeper and more meaningful learning. When the learners are involved and interested in meaningful tasks, they learn more effectively, and are more likely to retain and transfer the information to other contexts (Kearsley & Schneiderman, 1998). In the context of online learning, student engagement is an
integrated learning process guided by a set of motivational beliefs, behaviors, and metacognitive tasks that are planned and adapted to support the pursuit of personal goals (Wang, 2017).

The multifaceted nature of engagement (Liu, Wang, & Tai, 2016; O’Brien & Toms, 2008; Trevino & Webster, 1992) is typically described as having three or four components (Ding, Kim, & Orey, 2017). The most prevalent conceptualization in the literature suggests that engagement consists of three distinct, yet interconnected elements: behavioral, emotional, and cognitive engagement (Fredricks, Blumenfeld, & Paris, 2004). In the current research, the four-component model proposed by Fredricks et al. (2016) is adopted to represent online learning engagement, including cognitive engagement, behavioral engagement, emotional engagement, and social engagement.

Based on the definition given by the previous researchers (e.g. Fredricks et al., 2004; Meece, Blumenfeld, & Hoyle, 1988; Finn, 1989; Finn and Zimmer, 2012), a summary of these four dimensions is provided as follows. Behavioral engagement refers to the participation, effort, attention, persistence, positive conduct, and the absence of disruptive behavior. Cognitive engagement focuses on the student’s level of investment in learning, including being thoughtful, strategic, and willing to exert the necessary effort for comprehension of complex ideas or mastery of difficult skills. Emotional engagement represents students’ psychological reactions to academic environments, such as boredom or enjoyment. Social engagement concerns students’ prosocial behavior in classrooms and the quality of interactions with peers around instructional content.

Considering the dramatic differences between conventional classroom-based and online learning environment, subject-specific measures of engagement need to be developed for assessing learners’ engagement within different academic domains in online settings. Following Fredricks et al. (2016) theoretical model, the current research aimed at examining English language learners’ engagement in an online learning environment. To achieve the research purpose, this study seeks to adapt and validate the math and science engagement scale of Fredricks et al. (2016) with the participation of Chinese EFL learners.

3. Method

3.1 Participants

This present study was undertaken in a compulsory and credit-bearing EFL language course at a comprehensive university in the northeast of China. In this study, a convenience sampling strategy was used with a total of 454 freshmen (338 males and 116 females; average age: 19.3 years) who took this online class during the pandemic lockdown period were involved in this study. All the participants voluntarily responded to the questionnaire anonymously in one setting. It took about 5-10 minutes for the participants to complete the questionnaires.

3.2 Instrument

The questionnaire was adapted from the previous instrument of Fredricks et al. (2016). The original survey was reported to have high overall internal consistency reliability (Wang et al., 2016). The following is a brief description of the questionnaire. First, the four factors in the original questionnaire for measuring Math and Science Engagement (Fredricks et al., 2016), namely, “Cognitive engagement,” “Behavioral engagement,” “emotional engagement,” and “Social engagement” were maintained in the OLLE scale, but the questionnaire items were changed slightly by replacing “learning math and science” with “learning English online.” For instance, we changed the item ‘I try to connect what I am learning to things I have learned before.’ to ‘I try to connect what I am learning in this online English class to things I have learned before.’ After modifying the previous questionnaire, three experts in English language education were invited to examine the content of all the questionnaire items, providing expert validity for the survey. Eight university students were invited to think aloud when completing the survey, and made further minor revisions of the wording of certain items based on their comments. All the questionnaire items were presented in students’ native language, Chinese, on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The sample items of the four subscales are presented below:
3.3 Research procedure

Both exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were conducted to address the construct validity of the OLLE scale. The Statistical Package for the Social Sciences (SPSS 22.0) and the Analysis of Moment Structures (AMOS 24.0) were utilized to fulfill the purpose of this study. The procedure of data analysis included the following phases. First of all, the quantitative data obtained from the 454 participants were randomly divided into equal halves with 227 students respectively. Second, exploratory factor analysis (EFA) was used to test the construct validity of the OLLE questionnaire among the first set of participants (227 students, including 163 male and 64 female students). Then, confirmatory factor analysis (CFA) was performed to provide the constructive validity of the scale among the second set of participants (227 students with another 175 male and 52 female students).

4. Results

4.1 Exploratory factor analyses

The scale for evaluating students’ online language learning engagement in this research was adapted from the Math and Science Engagement Scale developed by Fredricks et al. (2016). Therefore, a re-examination of the factor structure and reliability of the factors for this survey was performed with the method of principal component factoring and Varimax rotation. Following the principle stated by Stevens (2012), items with loadings weighted greater than 0.40 on the relevant factor and less than 0.40 on all the other factors were kept in the finalized OLLE.

Table 1 shows the results of the exploratory factor analysis for the OLLE instrument. As a result, a total of 19 items were retained and grouped into four factors. The four factors were “cognitive engagement” (α= 0.90, Mean =3.20, SD = 0.98), “behavioral engagement’ (α= 0.87, Mean =3.48, SD = 0.87), “emotional engagement” (α= 0.88, Mean =2.90, SD = 0.98), and “social engagement” (α= 0.90, Mean =3.33, SD = 0.92). The factor loadings for all the items were all greater than 0.60, ranging from 0.68 to 0.86. The total variance explained for the scale was 69.24%. The Cronbach’s alpha value for each factor ranged from 0.87 to 0.90, indicating satisfactory internal reliability and validity for conducting further confirmatory factor analysis.

Table 1. The EFA analysis of the OLLE scale (N = 227)

<table>
<thead>
<tr>
<th></th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
<th>Factor 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cognitive Engagement (CE)</strong> (M=3.20 , SD=0.98, α=0.90)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive 1</td>
<td>0.76</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive 2</td>
<td>0.72</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive 3</td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive 4</td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive 5</td>
<td>0.74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive 6</td>
<td>0.71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Behavioral Engagement (BE)</strong> (M=3.48 , SD=0.87, α=0.87)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavioral 1</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavioral 2</td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavioral 3</td>
<td>0.79</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavioral 4</td>
<td>0.68</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In order to verify the construct of the OLLE scale, confirmatory factors analysis was conducted. As shown in Table 2, all Average Variance Extracted values (AVE) of components of online learning engagement had exceeded the cut-off value of 0.50, the Composite Reliability values (CR) ranged from 0.88 to 0.91, all alpha values were above 0.87 and the overall Cronbach’s value was 0.93. Therefore, the reliability of the questionnaire was established. Moreover, $\chi^2$/df = 2.138, RMSEA = 0.072, IFI = 0.93, CFI = 0.93, NFI = 0.88, GFI = 0.85. According to Hair et al. (2006), a reasonable value of the ratio of chi-square to the degree of freedom ($\chi^2$/df) is under 3.0. A reasonable fit of Root Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR) are acceptable if below 0.08. As suggested by Hair et al. (2006) and Gefen, Straub, and Boudreau (2000), the Incremental Fit Index (IFI), the Comparative Fit Index (CFI), the Normed-Fit Index (NFI), and the Goodness of Fit Index (GFI), are best if above 0.90 and demonstrate marginal acceptance if above 0.80. Hence, statistics all indicated that OLLE scale demonstrates good reliability and structural validity.

Table 2. The CFA analysis of the OLLE scale ($N = 227$)
### 5. Discussion

First and foremost, this paper reports on the development and validation analyses for the scale of students’ online language learning engagement conducted in an EFL online learning context in China. Through both exploratory and confirmatory factor analyses, the findings indicated that learners’ online language learning engagement include four factors, named as cognitive engagement, behavioral engagement, emotional engagement, and social engagement, revealing similar results as those of Fredricks et al.’s (2016) study. However, different from the previous studies conducted among primary school students, this research focused on the EFL learners in tertiary education. The findings presents the significance to duplicate the validation processes within different educational levels and subject matters to confirm valid clarification of the survey results.

The results of this study also revealed that EFL learners demonstrated the highest level of behavioral engagement followed by social engagement and cognitive engagement, while they had the lowest scores in emotional engagement. This finding implied that EFL learners believed that although they actively involved in task-based online activities of English language learning, they still displayed inadequate interest or positive emotional reactions to the online learning context. This finding is in conjunction with Xie, Debacker, and Ferguson’s (2006) finding in which students’ perceived enjoyment in online learning activities descended steadily throughout the semester. One of the possible reasons for this might result from lacking opportunities to interact, collaborate and receive feedback and social support in distance education (Tuckman, 2007). Similar results were also obtained in a study conducted in a K-12 online learning setting by Kim, Park and Cozart (2014). In their study, they found that lack of interpersonal interactions and the chance of receiving social support from peers may give rise to students’ anger that impeded their pace and actions of studying.

This study has several limitations and consequently implications for future research. First, it should be cautioned that the current study only employed quantitative measures. Future study could also combine other data collection methods, such as interview, performance-based method, in order to examine students’ OLLE thoroughly. Second, the diversity of research sample needs to be enriched. The use of more ethnically and socioeconomically diverse samples is important for understanding whether some dimensions of engagement are more important than others for enhancing online language learning engagement among minority youths and students from low-income families (Wang, Fredricks, Ye, Hofkens, & Linn, 2016). Finally, as students’ engagement in online learning activities may change over time (Liu, Chen, Lin, & Huang, 2017), the OLLE instrument needs to be refined and updated, which would be critical for understanding student engagement in different contexts, and designing pedagogical interventions aimed at increasing student engagement in a long-term online learning experience.

### Acknowledgements

This research was supported by “The Fundamental Research Funds for the Central Universities” [2020FZZL01], and the [International Joint Research Project of Faculty of Education, Beijing Normal University] under Grant [ICER201902].

<table>
<thead>
<tr>
<th>Social Engagement</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Social 1</td>
<td>0.72</td>
<td>–</td>
<td>0.91</td>
<td>0.63</td>
<td>0.91</td>
<td>3.26</td>
<td>0.95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social 2</td>
<td>0.83</td>
<td>11.83*</td>
<td>3.43</td>
<td>1.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social 3</td>
<td>0.84</td>
<td>11.99*</td>
<td>3.13</td>
<td>1.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social 4</td>
<td>0.86</td>
<td>12.26*</td>
<td>3.00</td>
<td>1.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total alpha: 0.93

Notes: CFA: Confirmatory Factor Analysis; CR; Composite Reliability; AVE: Average Variance Extracted; SD: Standard deviation.
References
Evaluation of a gamified augmented reality mobile app to support English language learning among non-native speakers

Dan WANG a, Mas Nida MD.KHAMBARI b*, Su Luan WONG c & Abu Bakar MOHAMED RAZALI d

Faculty of Educational Studies, Universiti Putra Malaysia, Malaysia *khamasnida@upm.edu.my

Abstract: This study proposes a gamified augmented reality mobile app (GARMA) to support the learning of English among college students in China. The app provides different scenes for students to learn English in a real environment. Also, it incorporates 3D images and videos supported by AR technology in each scene. Furthermore, an AR map and a leaderboard are adopted to create a gamified learning environment. The objectives of this preliminary study are to measure the performance and acceptance of the app in English language learning. In this study, the app was tested by a group of college students, English teachers and AR technicians through a survey. Based on the findings of the survey, further interviews were conducted. The results of the study show the app has good performance. Moreover, the study also reveals that the college students and teachers have a good acceptance of the app.

Keywords: Mobile augmented reality, gamification, English learning, app performance, user acceptance, vocational college

1. Introduction

In China, due to the examination-driven English teaching trend, the teaching of contents overemphasizes on book knowledge and overlooks the connection between knowledge and students’ daily experience, resulting in a dull learning atmosphere (Liu, 2011). In addition to this, another study found that students are rarely given the chance for oral English communication practice (Deng, 2019). As students have few opportunities to engage themselves in orally practicing English in class, they gradually lost interest in learning English, and eventually struggle to respond to the teaching and learning process (Liao, 2012). Currently, different kinds of teaching reforms have been implemented in English classes in China such as gamification and internet-based learning methods (Yang, 2010; Min, 2015; Wang, 2007). Although the initiatives have brought some changes to the teaching form, they have not effectively solved the problem of students’ lack of interest and engagement in learning English (Deng, 2019). How to cultivate students’ learning interest and improve their English language ability remains a question for English teachers to ponder over in the actual English language teaching. With the further development of the technology, emerging technologies such as augmented reality (AR) technology which has evident advantages in enhancing learning interest and engagement are bringing a new round of revolution in teaching (Goldman Sachs, 2016). AR technology enables students to interact with the real environment which is considered as an impossible leaning approach before (Billinghurst, 2002). Therefore, the researchers proposed a gamified augmented reality mobile app to support the learning of English. Meanwhile, this study used a pilot study to measure the performance of the app in terms of markers, speed, interface and leaderboard and evaluate the acceptance of the app in the light of perceived usefulness, perceived ease of use and perceived enjoyment.
2. Literature review

Gamification has the potential to boost learners’ motivation, engagement and performance through creating a comparative learning circumstance (Kim, Rothrock & Freivalds, 2016), combining numerous different elements such as points, badges, leader boards, rewards, feedback, story, levels and challenge to gamify the learning setting (Hamari, Koivisto & Sarsa, 2014). Integrating the gamified teaching method into English language teaching enables students to be more participative and practice English in communication (Yang 2010), as well as improve the situation of dull learning atmosphere and monotonous content (Min, 2015). Other than employing gamification as a method for teaching, AR has been used as a tool to enhance learner’s motivation by combining different forms of virtual materials such as virtual figures, vivid animation and sound to construct a real and virtual learning environment (Liu, Tan & Chu, 2010; Solak & Cakir, 2015). In English language learning, AR technology can enhance students’ learning interest and effectively combine with gamification to construct an attractive learning environment (Wang & Md. Khambari, 2020). Besides, some studies evaluated the applications of AR in educational areas and have proven that the applications of AR acquired good performance and acceptance including perceived usefulness, perceived ease of use and perceived enjoyment (Monroy Reyes, Vergara Villegas, Miranda Bojórquez, Cruz Sánchez & Nandayapa, 2016; Haugstvedt & Krogstie, 2012; Di Serio, Ibáñez & Kloos, 2013). This study, therefore, aims to measure the performance and acceptance of an English learning apps with AR and gamification integration in a higher education institution in China.

3. The gamified augmented reality mobile app design

3.1 The architecture of the app

The gamified augmented reality mobile app, using Ali elastic compute service and object storage service, was developed by the researchers using unity 3D software. The elastic compute service (ECS) provides a virtual server that implements a function of a leaderboard and object storage service (OSS) offers a storage function for videos. The use of the virtual server decreases the running cost without the need to buy hardware and it is flexible and convenient for a technician or a teacher to operate through remote control (Rao, Sasidhar & Kumar, 2012). The app, detecting markers located on real objects, can be used to learn English with corresponding images, videos and 3D in different real scenes. Based on the functions of the app, the whole leaning process can be flexibly designed to be a gamified AR learning process. Each learning process was designed to immerse students in a real scene to learn English and improve their writing, speaking, reading and listening skills.

3.2 A lesson conducted for the gamified augmented reality mobile app

At the beginning of English courses, students will get an AR map including many learning scenes such as supermarket, library, clinic and gym and the AR map is regarded as a course trigger marker. Then each learning group uses the app to scan the first scene they have chosen in the AR map to acquire an order of learning scenes. In this study, a supermarket scene was selected to apply the app and the five students work as a learning group. They only need to learn in the supermarket scene. After the students get paper reference materials which assist them learn English in the scene, the learning group will begin to learn English. The students will be brought into a scene of a supermarket where they will first scan the picture named AR English class to obtain a video regarding learning requirements in the supermarket scene and scan a real ten yuan to obtain a relevant English introduction regarding the supermarket and 3D snacks will be displayed. Then, the students go to shelves to scan on the snacks. Each snacks that they have scanned will be shown an AR of the elaborations on the flavor and features, which are the seed materials, namely AR videos, AR images, and real physical environment. With the seed materials, the students need to write a 100-word English essay related to the actual scenario. Then, every group must video record themselves on GARMA reading the short essay. When done, they have to click “Submit”. Immediately, the app generates a timely ranking. If a group is not satisfied with the current writing and recording and wants to revise them again, the group can modify and submit them
again. The learning time will be recalculated. Materials including the iPad, the paper reference materials and essays need to be submitted to the course instructor after class for each group. When the class is over, the students can revisit their video recording by scanning a trigger image. This feature allows the students share their assignments in different scenes and different teams, which may facilitate students’ discussion after class.

4. Research Methodology

This study adopts an exploratory survey method to measure the performance and acceptance of the gamified augmented reality mobile app. This performance evaluation mainly focused on the app’s characteristics including marker system, speed and interface as claimed by Monroy Reyes, Vergara Villegas, Miranda Bojórquez, Cruz Sánchez and Nandayapa (2016). The questionnaire, using a Likert scale ranging from 1 (very poor) to 5 (very good) was also adopted from their work. Furthermore, because the app was also characterized by gamification element, namely leaderboard, thus, it is included in performance evaluation. Additionally, in the evaluation of the acceptance of the app, there are three dimensions in the questionnaire. They were perceived usefulness, perceived ease of use and perceived enjoyment as suggested by Haugstvedt and Krogstie (2012) who claimed that the three dimensions were significant determinants for the acceptance of AR application. In the questionnaire, the items of perceived usefulness were based on the work of Hsieh, Kuo and Lin (2014) whose study focused on English discipline, which was more suitable for the evaluation of usefulness in the study. The questionnaire used a Likert scale which was a seven-point semantic differential. Besides, based on the results of the survey, the researchers implemented interviews with the participants to explore deeply the performance and acceptance of the app. Five technicians who have more than five years of experience in developing AR app participated in the initial discussion of GARMA development and knew the basic architecture of the app. Therefore, they were selected to evaluate the app. As for practitioners, five English teachers and five college students, whom have experience in using mobile app, were recruited to participate in the evaluation. These participants were introduced and briefed about GARMA before evaluation were carried out.

5. Results and discussion

The performance of the GARMA was evaluated by the five technicians through a questionnaire. The findings of the survey are presented in Table 1.

Table 1. The Results of the Questionnaire Used to Evaluate the Performance of the App

<table>
<thead>
<tr>
<th>Items</th>
<th>Percentage (%)</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Very poor</td>
<td>Poor</td>
</tr>
<tr>
<td>1. Markers (detection)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2. Speed (app)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3. Speed (object storage service)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4. Leaderboard (virtual server)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5. Interface</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

According to Table 1, it showed that the total average mark was over 4.00. All of the items except item 3 (Speed of object storage service) obtained over 4.00 mean value. Item 4 (Leaderboard) had the highest mean value (M=4.40). However, item 3 (Speed of object storage service) had the lowest mean score (M=3.80). The results showed that in general, the performance of the gamified augmented
reality mobile app acquired a good feedback from the technicians. The feature of the leaderboard was greatly approved by the technicians. However, the performance of the speed of object storage service just reached a standard performance. The object storage service is related to the video display and it will impact the corresponding videos to superimpose on the markers while students scanning the markers. In the further interviews, some of the technician participants suggested that (i) the AR videos’ response speed need to be upgraded, (ii) a bigger cloud storage service is needed for the objects, and (iii) a more structured graphical user interface could improve the mobile app usability among college students. However, it is worth noting that addressing both recommendations will incur more costs to the researchers.

The findings regarding the acceptance of the gamified augmented reality mobile app were collected from five college teachers and five students and displayed in Table 2.

Table 2. The Findings of the Questionnaire Used to Evaluate the Acceptance of the App

<table>
<thead>
<tr>
<th>Items (Perceived Enjoyment)</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>PE1 Disgusting-enjoyable.</td>
<td>6</td>
<td>7</td>
<td>6.50</td>
</tr>
<tr>
<td>PE2 Unpleased-pleasant.</td>
<td>6</td>
<td>7</td>
<td>6.40</td>
</tr>
<tr>
<td>PE3 Dull-exciting.</td>
<td>6</td>
<td>7</td>
<td>6.50</td>
</tr>
<tr>
<td>PE4 Boring-interesting.</td>
<td>6</td>
<td>7</td>
<td>6.60</td>
</tr>
<tr>
<td>Total mean</td>
<td></td>
<td></td>
<td><em>6.50</em></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Items (Perceived Ease of Use)</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>PEU1 Interaction with the app is clear and understandable.</td>
<td>6</td>
<td>7</td>
<td>6.70</td>
</tr>
<tr>
<td>PEU2 Interaction with the app does not require a lot mental effort.</td>
<td>6</td>
<td>7</td>
<td>6.70</td>
</tr>
<tr>
<td>PEU3 I find the app easy to use.</td>
<td>5</td>
<td>7</td>
<td>6.30</td>
</tr>
<tr>
<td>PEU4 I find it easy to get the app to do what I want it to do.</td>
<td>5</td>
<td>7</td>
<td>6.20</td>
</tr>
<tr>
<td>Total mean</td>
<td></td>
<td></td>
<td><em>6.48</em></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Items (Perceived Usefulness)</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>PU1 The app can enrich the learning contents.</td>
<td>6</td>
<td>7</td>
<td>6.50</td>
</tr>
<tr>
<td>PU2 The instruction of the app is so clear that I understand the learning contents effectively.</td>
<td>6</td>
<td>7</td>
<td>6.50</td>
</tr>
<tr>
<td>PU3 The instruction provided by the app is easy to understand and follow.</td>
<td>6</td>
<td>7</td>
<td>6.60</td>
</tr>
<tr>
<td>PU4 The app is helpful in my learning.</td>
<td>6</td>
<td>7</td>
<td>6.70</td>
</tr>
<tr>
<td>PU5 The app can help me learn better.</td>
<td>6</td>
<td>7</td>
<td>6.60</td>
</tr>
<tr>
<td>PU6 Generally speaking, I find out the app is useful in my learning.</td>
<td>6</td>
<td>7</td>
<td>6.50</td>
</tr>
<tr>
<td>Total mean</td>
<td></td>
<td></td>
<td><em>6.57</em></td>
</tr>
</tbody>
</table>

Total mean for the acceptance of the app *6.52*

According to Table 2, the app was well rated with an average score of 6.52, which showed that the teachers and students generally had a good acceptance of the app. Moreover, the perceived enjoyment achieved a high mean value (M= 6.50). All four items obtained high mean values and each score of the all items was 6 or 7. Item PE4 (Boring-interesting) was well rated with the highest mean value (M= 6.60). The findings showed that the students and teachers enjoyed the learning process and felt interesting in learning English. Based on the results, the researchers conducted further interviews. The students claimed that the function of the leaderboard was interesting and it provided a sense of competition, which motivated them to learn English. This finding is congruent with Yang, Quadir and Chen (2016) who revealed that leaderboard ranking had positive influence on students’ English learning. Besides, the participants also stated that learning English in real environment also made them feel interesting in the whole learning process through scanning the real objects and interacting with the real learning environment. The finding is confirmed by Hsieh, Kuo and Lin (2014) who claimed that AR environment cloud enhance students’ enthusiasm and interest in learning English.

Furthermore, according to Table 2, the ease of use of the app also acquired a good rating with the total mean value of 6.48. The minimum score of the four items was 5 and the mean values of the four items ranged between 6.20 and 6.70. Item PEU1 (Interaction with the app is clear and understandable) and item PEU2 (Interaction with the app does not require a lot mental effort) both obtained high mean values (M=6.7). However, item PEU3 (I find the app easy to use) and item PEU 4 (I find it easy to get

513
the app to do what I want it to do) were relatively low with the mean values of 6.30 and 6.20 respectively. Overall, the results indicated that the ease of use of the app was good. Hsieh’s study (2016) reported that AR was easy to use echoed the finding. However, the findings also showed that there were still some issues in the process. Based on the findings, participants are interviewed and reported that the learning process was easier to control and understand than they expected. They thought that the interface was helpful for them to understand how to operate the app. Also, they considered that the marker detection was easy to control and the respond speed is good. But one student pointed out that the video loading needed to be improved to enhance the ease of use.

Additionally, according to Table 2, the usefulness of the app was well rated with a mean score of 6.57. All the six items had a mean value over 6.50 and item PU4 (The app is helpful in my leaning) achieved a highest mean score (M=6.70). The results indicated that the participants considered that the app was useful for their English language learning. Moreover, the further interviews showed that the participants thought that the app provided more opportunities and spaces for learners to practice English in whole learning process, which was useful to improve English writing, speaking, reading and listening skills. Besides, the findings of the interviews also pointed out that interacting with the real learning environment was useful to improve their English abilities, especially the speaking skill. This finding is supported by Liu (2009) who claimed that real-life situation can effectively improve speaking skill. The excerpts of evidences of interview transcripts:

Zhang (student): “Compared to traditional English class, this learning method lets me interact with items in a real supermarket and scan the real money to acquire the learning content, which motivated me to learn English, and the process is very interesting. Besides, the leaderboard is also interesting and I become more likely to communicate in English in this environment”.

Zhou (student): “the interface is easy to understand and the app is easier to use than I imaged before. But if the video loading can be faster, it should be better”.

Li (teacher): “In whole learning process, compared to the traditional English class, this app provides many opportunities and spaces for students to practice English including writing, speaking, reading and listening skills, which is useful for improving English abilities”.

Based on the results offered by the participants, the technical experts gave positive comments regarding the performance evaluation of the gamified augmented reality mobile app. Also, they confirmed that its performance can effectively support the implementation of various teaching aims. Besides, the app was considered to be a useful and interesting app for learning English and acquired good acceptance. Although GARMA was highly rated by all the participants, many parts still need to be further refined and optimized.

6. Conclusion

GARMA is developed to provide an immersive English learning environment by incorporating AR technology and gamification elements in a real scene. This app may be helpful in immersing students in a real location to learn English, providing a more flexible and interesting learning process that is relatable to daily encounters. The findings of this study suggest that GARMA has good performance in the learning process and also obtains good acceptance in English language learning. However, the app also needs to be further improved, especially in the object storage service. Although the sample size is small, it is worth to note that the aim of this study is to improve the app, rather than generalizing the findings to a larger population. The suggestions collected from the participants are essential in improving GARMA so that it can be utilized as a new learning approach to learn English in several other different scenes. The researchers will also continue to study the effect of GARMA in English language learning and explore deeply on the learning process through the gamified AR environment.

Acknowledgements

We would like to thank the Centre for Academic Development, Universiti Putra Malaysia, for funding this research via the Incentive for Teaching and Learning Grant.
References


A New Technology Design for Personalized Incidental Vocabulary Learning using Lifelog Image Analysis

Mohammad Nehal HASNINEa*, Kousuke MOURIb, Gokhan AKCHAPINARc, Mahmoud Mohamed Hussien AHMEDd & Hiroshi UEDAA

aResearch Center for Computing and Multimedia Studies, Hosei University, Japan
bAcademic Center for Computing and Media Studies, Kyoto University, Japan
cDepartment of Computer Education & Instructional Technology, Hacettepe University, Turkey
dDepartment of Educational Technology, South Valley University, Egypt
*nehal.hasnine.79@hosei.ac.jp

Abstract: Incidental vocabulary learning is the process of learning foreign vocabulary without the intention of doing so. In language learning pedagogy, incidental learning is considered to be an effective way to enhance foreign vocabulary from context. Conventional vocabulary learning systems in ubiquitous learning scenarios are developed primarily for supporting intentional vocabulary learning. When learning foreign vocabulary using conventional tools, it is not feasible to learn incidental vocabulary. Moreover, there does not exist a general framework to formulate lifelong images along with typical ubiquitous learning logs such as location, time, and demographics. Therefore, this research precisely looked into the scope of formulating lifelog images as sensor data for enhancing incidental vocabulary knowledge of English as Foreign Language (EFL) learners. The main pursuit of this article is to introduce a model for enriching ubiquitous learning literature, in which incidental vocabulary can be learned. In this study, a technology-enhance environment for incidental vocabulary learning in the EFL context is presented. The research objectives are to- i) design a new technology that is capable of analyzing lifelong images in order to generate a bag of incidental vocabularies, ii) determining the top-5 vocabularies that could be recommended to the learner, and iii) to automatically create learning material for each of the recommended words. We employed visual content analytics on lifelong learning images using object detection method. This is known to be an applied artificial intelligence method that is applicable across a variety of fields, such as natural language processing, computer vision, and others. We aimed to apply this applied AI method to design a new technology-enhanced learning environment.

Keywords: AI in language learning, applied AI, incidental vocabulary learning, informal learning, lifelog image analysis, technology-enhance language learning

1. Introduction

For language learners, vocabulary learning is an indispensable activity to achieve competence in the target language. Vocabulary enriches learners’ integrated language skills such as listening, speaking, reading; which later on facilitates fluent conversation and effective writing (Ahmad, 2012). In language learning pedagogy, several theories and approaches have tried to account for the specific way that learners’ learning takes place. It seems, however, that vocabulary learning depends on the type of cognitive process in which the learner is engaged (Ramos & Dario, 2015). Two commonly discussed pedagogical methods that are directly associated with learners’ cognitive processes are intentional learning and incidental learning. Intentional learning is defined as being designed, planned for, or intended by teachers or students (Yali, 2010). In contrast, incidental learning is the process of learning something without the intention of doing so (Ahmad, 2012). It is also learning one thing while intending to learn another (Richards & Schmidt, 2013). This type of learning strategies can be used to enhance vocabulary swiftly. Research indicates that, in comparison with the intentional vocabulary learning method, the incidental learning method can be effective. According to (Ahmad, 2012), intentional
Vocabulary learning often learned based on synonyms, antonyms, word substitution, multiple-choice, scrambled words, and crossword puzzles, regardless of context, is not so effective, because learners are more prone to rote learning. Hence, research emphasis needs to be given on incidental vocabulary learning. Lifelog images are those often captured to record learners’ learning experiences. Other research fields addressed lifelog images as visual lifelogging as they are logged by wearable devices such as GoPro, MeCam, Looxcie, or Google Glass. Without arguing, each of the lifelog images has a story to tell for the capturer, hence, they are a rich source of data. Research already evident that, lifelog images are the source of valuable information because the pictures taken offer considerable potential for knowledge mining concerning how people live their lives; hence, they open up new opportunities for many potential applications in fields including healthcare, security, leisure, and the quantified self (Bolaños et al., 2017). With the convenience of smartphones and wearable technologies, high temporal resolution, and are more suitable to record specific moments from language learners, such as cultural experience, authentic learning experience, problem-based learning, task-based learning experiences, etc. Therefore, language learners’ lifelog images together with other ubiquitous logs need to be analyzed to improve language learning particularly vocabulary learning.

To our best knowledge, as yet, there does not exist any framework that formulates lifelong images along with typical ubiquitous learning logs such as location, time, and user information. Therefore, we aim to design a new technology in technology-enhanced language learning with the objective to support incidental vocabulary learning from lifelogging. With this study, we aim to overcome one criticism about conventional ubiquitous learning systems namely the scope of learning new vocabulary is limited.

2. Literature Review

2.1 Previous works by this research team

Researching about the incidental vocabulary learning using ubiquitous learning log and other sensor data is our ongoing project. Some recent outcomes that related to this work are- in 2019, we introduced a location-based word recommendation system for incidental vocabulary learning in English as Foreign Language (EFL) context (Mohammad Nehal Hasnine et al., 2019). This system uses EFL learners’ vocabulary learning histories and location data as the input; then it leverages association rule mining for discovering new knowledge and non-trivial patterns hidden in learning history; finally generates words that are associated to that particular location. In 2019, our other study used images’ visual contents as contextual clues to generate special learning contexts where new vocabularies believed to be learned from sentences (Hasnine, Flanagan, et al., 2019). To develop this technology, we leveraged the power of automatic image captioning. Later on, in ICCE 2019, we introduced a model that detects the textual information from lifelog images (Mouri et al., 2019). This model relies on several APIs to detect objects and textual clues from images. These clues were used to support foreign language learning by bridging cyber space learning with physical learning.

2.2 Related works by global researchers

In order to support intentional vocabulary learning using ubiquitous and computer-aided technologies, SCROLL, UEVL, MALL, and U-Arabic systems are commonly cited. These tools use location data, context information, word learning histories etc. to support intentional learning. However, these tools do not support incidental learning at all; the idea of learning experience sharing is noticed, though. For incidental vocabulary learning, study by (Song & Fox, 2008) used personal digital assistant as the learning technology to teach English words to university students. A one-year-long case studies on undergraduate students revealed that personal device assistants can be used in more flexible, novel and extended ways for English as a Foreign Language (EFL) vocabulary teaching and learning in higher education, taking student needs and contexts into consideration. Besides, video captioning has been tested for incidental vocabulary learning by (Montero Perez et al., 2014). This study suggests that captioning did not affect much on comprehension nor meaning recall. Electronic dictionaries as the tool are used in the early 2000. Study conducted by (Laufer, 2000) revealed that students who learned by electronic text performed significantly better than the paper text group in learning low frequency words.
Our meta-analysis on previous studies indicates that, by far, time, place, contextual memos are used as contextual clues to support vocabulary learning. However, lifelog images as contextual clues is overlooked. Recently, wearable cameras on the market like GoPro, MeCam, Looxcie, or Google Glass are relatively getting popularity for high temporal resolution and are more suitable to record specific moments, such as cultural experiences. Therefore, lifelog images can be used as contextual clues to support foreign language learning. Consequently, we propose this study where foreign language learners lifelog image are analyzed to support incidental vocabulary learning. This, we believe, will bring new dimension in the design of technology-enhance language learning using applied AI.

3. Technology Design

3.1 Overview of the Model

We begin by clearly introducing the objectives of this study. This technology is designed for foreign language learners to support their incidental vocabulary learning. Therefore, this technology aims to broaden learners’ opportunities to engage with new vocabulary, create learning materials, and reflect on their memory. Unlike existing ubiquitous language learning systems where place, time, and handwritten memos are used the contextual clues; the newness of this proposed technology is the usage of lifelogging images as the primary contextual clue. We leveraged lifelog images because the visual contents of lifelogging contain powerful social interaction, the analysis of social interactions in lifelogging data is of fundamental importance to understanding human behavior; and the presence of people and social interactions are consistently associated with our memory. Hence, we hypothesized that as lifelog images contain powerful information, they need to be analyzed for educational uses (Bolaños et al., 2017). We introduce the design of the model in Figure 1. Later in the 3.3, we discuss each of the key components of the model.

![Figure 1. The Design of the Model](image)

3.2 Methods and Materials

The proposed model begins by reading data from three modular data sources. At present, there are three main sources for reading lifelog images. The first source is, the database of the AIVAS (Appropriate Image-based Vocabulary Learning System) system; the second is a dataset; the third is manual uploading. After the data is read, the image analysis task is begun. This task is done by two AI-based models namely an image captioning model and a cognitive vision model. The image captioning model
generates three possible contexts in the form of sentences; whereas the cognitive vision model produces a bag of words. The bag of words contains all of the possible vocabularies that the learner could learn as incidental learning. After that, the model determines top-5 incidental vocabularies based on the word’s confidence level. That is, five words with the highest confidence are recommended for learning. Finally, for each of the recommended words, learning material is created automatically. For the automatic creation of the learning materials, we used our previously developed system (M. N. Hasnine et al., 2016; Mohammad Nehal Hasnine, 2018; Mohammad Nehal Hasnine et al., 2017).

3.3 Technical Specifications of the Components

The proposed design heavily relies on external APIs and an online language learning tool. Here, we provide a brief introduction to the technologies that are used to design this new technology.

- AIVAS system: It is a web-based language learning tool that assists learners in creating learning materials (M. N. Hasnine et al., 2016; Mohammad Nehal Hasnine et al., 2015).
- Lifelog image dataset: This is a dataset that contains about 10000 lifelog images in total. These images were captured by foreign language learners using lifelog camera and smartphone camera technology. A mobile technology called SCROLL (System for Capturing and Reminding of Learning Logs) is used to capture the images. The dataset (Ogata et al., 2018) is available for research purposes.
- Camera album and local directory: Our model can read image data from a local computer and camera album of a smartphone. This is for the manual testing.
- MAX Image Caption Generation API: This is an API developed by IBM. This API uses artificial intelligence for understanding scenes and uses NLP methods for describing scenes in natural language. The model we used for our study is based on Show and Tell image-to-text generation model.
- Microsoft Cognitive Vision API: This API is developed by Microsoft used for extracting rich information from RGB images. This API uses Microsoft’s computer vision services for analyzing image scenes. The input of this API is a natural image. The output is a bag of vocabulary. Each of the vocabularies listed in the bag is represented with a tag, object name, description, and category of the object.
- Microsoft Translation API: This is a cloud-based service for machine translation offered by Microsoft. This API takes a word or a phrase as input and returns the corresponding translation of the target language that the user wishes to get.
- VoiceRSS API: This is a text to speech API. This API read textual content for converting to speech. This is a free API that can be used for research purposes.
- AIVAS-AIRS: AIRS stands for Appropriate Image Recommendation System. In AIVAS, we use AIRS system for determining appropriate images for representing a word. This system uses the image ranking method (Mohammad Nehal Hasnine, 2018) for recommending appropriate images. This system is capable of ranking images from a set of corresponding images for a word that is downloaded from image search engine.

4. System Development

Based on the proposed model (described in Section 3), we developed a prototype of the incidental vocabulary learning system for supporting foreign language learners. At present, the system produces English vocabularies from lifelogging images. The system captures conventional ubiquitous logs such as time, place, intentional vocabulary together with a bag of incidental vocabulary. Figure 2 shows the functions to analyze lifelog image scenes for producing top-5 incidental vocabulary. The top-5 incidental vocabulary is determined by the model’s confidence for each of the vocabulary in the bag.

---

1 https://github.com/IBM/MAX-Image-Caption-Generator
2 https://github.com/tensorflow/models/tree/master/research/im2txt
4 http://www.voicerss.org/api/documentation.aspx
Figure 2. Incidental Vocabulary Generation from a Lifelog Image

Figure 3 is the interface where incidental vocabularies are recommended and those are studied by the learners. In this interface, a learning material can be created automatically for each vocabulary. Upon clicking on a word, a learning material will automatically be created and displayed on the interface. For creating a learning material, we used the learning material creation module of the AIVAS system (Mohammad Nehal Hasnine, 2018; Mohammad Nehal Hasnine et al., 2017).

Figure 3. An Automatically-generated Material for Learning for a New Word
5. Conclusion

Incidental vocabulary learning is an important aspect of language learning pedagogy. Yet, it is not feasible to learn incidental vocabulary using conventional learning tools. On top of that, research on conventional ubiquitous learning overlooked the power of visual lifelogging as the potential learning logs. It seems traditional definitions of ubiquitous learning is much closely associated with learning by logging and contextualization. Therefore, in this paper, we introduced a technology design for incidental vocabulary learning in the EFL context. As smartphone camera technologies along with other wearable technologies such as GoPro, MeCam, Looxcie, or Google Glass are becoming mainstream research topics, our proposed model would be a good fit for it.

In this paper, we introduced a new design for technology-enhanced language learning that aims to support incidental vocabulary learning in the EFL context. We first implemented the model in our local server. Then, we developed a prototype of the system. Interface displayed in Figure 2 is to analyze the model’s output with regards to top-5 objects to learn from a lifelogging image. The interface displayed in Figure 3 is for generating learning material for each word. The second interface can be used for memorizing new words and taking quizzes for reflection. With this technology, our goal is to give a learner scope to encounter with new vocabularies and learn them using multimedia annotations.

References

Exploring Online Reading Attention and Question-Generation Task Completion of English Learners at Different Proficiency Levels by Eye Tracking Method

Fu-Yun YUA, Wen-Wen CHENG* & Jon-Fan HUB

*a Institute of Education, National Cheng Kung University, Taiwan
b Department of Psychology, National Cheng Kung University, Taiwan
*wliebe@gmail.com

Abstract: This study investigated online reading attention from the viewing patterns of English learners at different proficiency levels as well as their completion of question-generation task on English grammar. Seven college students participated and were classified to low (n=3) and high (n=4) English proficiency. Eye tracker Tobii X60 was used to record data on the participants’ eye fixations within areas of interest (AOIs) and their visits between AOIs while reading online materials of different instructional functions (i.e., Review, Instruction, Situation, Related hints, and Unrelated hints) for student-generated questions task. Two major findings were obtained. First, there were salient different online reading attention patterns in terms of fixation duration and visit counts between students at different English proficiency levels. Specifically, based on Kruskal-Wallis tests, the High English Proficiency Group (HEPG) paid more attention to areas related to the targeted SGQ task, by visiting significantly more frequently on ‘Instruction’ and ‘Related hints’ type of information, and spending significantly more time attending to ‘Related’ useful information, rather than on ‘Unrelated’ information, as compared to the Low English Proficiency Group (LEPG), who seemed to have less attention tactics to help them filter out unrelated information. Furthermore, based on the Fisher exact tests, significant differences were detected between HEPG and LEPG on SGQ task completion, with HEPG having a higher task completion rate than LEPG. Limitations of this study are noted, and suggestions for instruction as well as future studies are provided.

Keywords: English grammar, eye tracking, individual differences, online reading attention, second language learning, task completion

1. Introduction

The crucial role of attention and its positive association with second language learning have been studied for decades (Al-Hejin, 2004; Dolgunsöz, 2015; Robinson, 1995; Robinson, Mackey, Gass, & Schmidt, 2012; Schmidt, 1990, 2001, 2010; Song, 1998). The Noticing Hypothesis has been proposed to accentuate that input does not become intake for language learning unless it is noticed (Izumi, 2002; Robinson, 1995; Schmidt, 1990, 2001). Researchers considering attention as a key factor affecting language learning examined its facilitative role for vocabulary acquisition (Godfroid, Ahn, Choi, Ballard, Cui, Johnston, et al., 2018; Godfroid, Boers, & Housen, 2013) and reading (Dolgunsöz, 2015; Godfroid et al., 2018) and have found empirical evidence substantiating the significant implications of attention and awareness for language learning (Bialystok & Feng, 2009; Chung & Segalowitz, 2004; Hama, & Leow, 2010; Gass, Svetics & Lemelin, 2003; Ishikawa, 2006; Robinson et al., 2012; Schmidt, 2012). Learners at different language proficiencies have later been found to develop different reading strategies and processing types. For instance, McCradden and Schraw’s research (2007) reported that proficient readers were aware of their information needs and, thus, had the ability to locate information and free themselves from irrelevant messages. Rouet and Britt (2011) found that when faced with several pieces of text, proficient readers can select information and make decisions for the most important, relevant, and accurate information. As shown in Yildiz and Çetinkaya’s findings
(2017), there is a significant relationship among good readers’ attention, reading speed, and comprehension.

While the effects of attention and the moderating role of language proficiency on language learning have been noted, existing studies mostly deal with learners processing reading materials for comprehension and vocabulary learning (i.e., the ‘knowledge consumption’ notion) without further calling for knowledge production. With the increasing recognition of the value of the learner-as-producer approach (Arruabarrena, Sánchez, Blanco, Vadillo, & Usandizaga, 2019), issues as to ‘if reading attention and task performance of English learners at different proficiency levels differ under the learner-generated content paradigm’ await to be investigated.

Today’s learners are born as digital natives and accustomed to learning from online materials (Prensky, 2001). Student-generated questions (SQG) as a pedagogically sound strategy to support knowledge construction and higher-order cognitive development on the part of learners has been well acknowledged (Song, 2016; Yu, 2009). In light of these, in this study, the researchers aimed to examine ‘if students at a high English proficiency level would better select relevant, helpful information when reading online short passages and complete the targeted SGQ task, as compared to students at a low English proficiency level.’ Two frequently used indices of eye tracking method (i.e., fixation duration and visit counts) are used for examining online reading attention from the viewing patterns of readers at different English proficiency levels for SGQ tasks. Fixation duration is the period of time when the reader’s gaze focuses still on an area of interest (AOI) for gaining information. A longer duration indicates a more complex task, a more difficult text, or readers have more interest toward the AOI (Kim, Dong, Xian, Upatising, & Yi, 2012). Visit is considered a good index for observing how readers switch their fixations among different stimulus features (Kim et al., 2012). It starts when a reader first focuses on an AOI and ends when he or she looks away from the AOI.

In summary, two research questions are examined in this study:

RQ#1. Are there significant differences in eye movement and online reading patterns in terms of fixation duration and visit counts of AOIs between learners in the High English Proficiency Group (HEPG) and Low English Proficiency Group (LEPG)?

RQ#2. Are there significant differences in SQG task completion between learners in HEPG and LEPG?

2. Methods

2.1 Participants

Seven students (2 males) from three universities in southern Taiwan participated in this study. Of these, three participants were classified as below B1 in the LEPG by a standardized test of the TOEIC based on the Common European Framework of Reference for Languages: Learning, Teaching Assessment (CEFR), with B1 level equivalent to the TOEIC score of 550. The other four with above B1 level were, thus, classified to HEPG.

2.2 Stimulus Materials, Study Context, and Equipment

The stimulus material produced by the researcher consists of three slides and was specified with different AOIs according to five instructional functions: Review, Instruction, Situation, Related hints, and Unrelated hints. For the study, the participants had to read the stimulus material for the SGQ task from a computer screen in an eye-tracking lab. Eye tracker Tobii X60 was used to collect data of the participants’ online viewing patterns, and Tobii Pro lab was used to record and calculate data.

For Slide 1 (see Figure 1), which is for ‘Review’ purposes, two examples related to the unreal second and third conditionals were presented for a quick summary of the two grammatical rules. For Slides 2 and 3, two cases (i.e., Cases 1 and 2) involving unreal conditionals and instruction for SGQ tasks were presented. For Case 1 (see the left of Figure 2), the general instruction for the reading and SGQ tasks (i.e., ‘Instruction’) was placed above a short passage (i.e., ‘Situation’), followed by the specific instruction for SGQ on the unreal third conditional (i.e., ‘Instruction’). The short passage
shown on the computer was from the intermediate level of the textbook, Top Notch 3 (Saslow & Asher, 2015). Placed at the bottom of Slide 2 were the same set of examples shown in Slide 1 (i.e., ‘Review’), with the unreal second conditional classified as ‘Unrelated hint’ and the unreal third conditional classified as the ‘Related hint,’ due to its direct relevancy to the SGQ task to be completed, which requested the participants to generate questions on the unreal third conditional with answers. For Slide 3 (the right of Figure 2), it showed Case 2 and had the same layout as in Case 1. The only difference was that a different reading passage from the intermediate level of Top Notch 3 was given on the computer as the ‘Situation’ for the reading and SGQ tasks.

![Figure 1](image1.png)

**Figure 1.** Stimulus material shown in Slide 1 and the AOI specified for data analysis

![Figure 2](image2.png)

**Figure 2.** Stimulus material shown in Slide 2 for Case 1 (left) and Slide 3 for Case 2 (right) and the AOs specified for data analysis

### 2.3 Data Compilation and Analysis of AOIs

In this study, fixation duration and visit counts of AOIs in Cases 1 and 2 were combined for the five instructional functions, respectively, before proceeding to data analyses. Non-parametric statistics of Kruskal-Wallis tests are adopted to test if there are significant differences between HEPG and LEPG in attention allocation via eye movement in terms of fixation duration and visit counts of AOIs.

### 2.4 Assessment and Data Analysis of SGQ Task Completion

The generated questions with answers for Cases 1 and 2 had to satisfy the criteria of ‘correct use of the unreal third conditional according to the given situation.’ Due to the small sample size in this study and the fact that the expected frequency is lower than five, non-parametric statistic of the Fisher exact test is applied to analyze if there are significant differences between HEPG and LEPG in SGQ task completion.

### 3. Results

With regard to RQ#1, as shown in Table 1, the mean scores of both indices (i.e., fixation duration and visit counts) for HEPG were all higher than those of LEPG except for the ‘Unrelated hints’ instructional function. Further Kruskal-Wallis tests showed that there were significant differences between HEPG and LEPG in ‘Review’ (fixation duration, \( p = .034 \)), ‘Instruction’ (visit counts, \( p = .05 \)), ‘Unrelated hints’ (visit counts, \( p = .032 \)), and ‘Related hints’ (fixation duration, \( p = .034 \); visit counts, \( p = .026 \)).

524
Table 1. Fixation Duration and Visit Counts of AOIs along the Five Instructional Functions of English Learners in HEPG and LEPG

<table>
<thead>
<tr>
<th>Group</th>
<th>Participants</th>
<th>Review</th>
<th>Instruction</th>
<th>Situation</th>
<th>Unrelated hints</th>
<th>Related hints</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Fixation Duration+</td>
<td>Fixation Duration</td>
<td>Visit Count</td>
<td>Fixation Duration</td>
<td>Visit Count</td>
</tr>
<tr>
<td>HEPG</td>
<td>P1</td>
<td>25.33</td>
<td>17.47</td>
<td>11</td>
<td>57.36</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>P2</td>
<td>14.04</td>
<td>27.1</td>
<td>23</td>
<td>91.25</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>P3</td>
<td>26.09</td>
<td>16.84</td>
<td>17</td>
<td>98.08</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>P4</td>
<td>21.27</td>
<td>12.64</td>
<td>18</td>
<td>54.19</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td>21.68</td>
<td>18.5</td>
<td>17.25</td>
<td>75.22</td>
<td>31.25</td>
</tr>
<tr>
<td>LEPG</td>
<td>P5</td>
<td>10.07</td>
<td>2.9</td>
<td>7</td>
<td>60.83</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>P6</td>
<td>11.38</td>
<td>15.02</td>
<td>9</td>
<td>53.24</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>P7</td>
<td>7.36</td>
<td>7.79</td>
<td>11</td>
<td>95.23</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>mean</td>
<td>9.60</td>
<td>8.57</td>
<td>9</td>
<td>69.77</td>
<td>28.67</td>
</tr>
</tbody>
</table>

Kruskal-Wallis tests

<table>
<thead>
<tr>
<th></th>
<th>4.50</th>
<th>3.13</th>
<th>3.85</th>
<th>.13</th>
<th>.29</th>
<th>1.13</th>
<th>4.58</th>
<th>4.50</th>
<th>4.94</th>
</tr>
</thead>
<tbody>
<tr>
<td>*p</td>
<td>.034*</td>
<td>.077</td>
<td>.05*</td>
<td>.724</td>
<td>.593</td>
<td>.032*</td>
<td>.289</td>
<td>.034*</td>
<td>.026*</td>
</tr>
</tbody>
</table>

*Only one area was specified in ‘Review’ so no visit counts metric was calculated, and only fixation duration metric was reported.
* < .05

With regard to RQ#2, as shown in Table 2, 87.5% of the participants in HEPG satisfactorily completed the SGQ task, but not a single participant in LEPG did. The Fisher exact test further found significant differences \((p=.005<.05)\) between HEPG and LEPG in SGQ task completion, with HEPG having a higher task completion rate than LEPG.

Table 2. SGQ Task Completion of English Learners in HEPG and LEPG

<table>
<thead>
<tr>
<th></th>
<th>No</th>
<th>Yes*</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEPC</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>HEPG</td>
<td>1</td>
<td>7</td>
</tr>
</tbody>
</table>

*Questions/answers correctly used the unreal third conditional for SGQ task

4. Discussion and Conclusion

To explore how learners at different English proficiency levels attended to online reading materials of different instructional functions for ‘student-generated content’ type of learning task, in this study SGQ was targeted, and eye tracking method was adopted where data on fixation duration and visit counts as well as SGQ task completion were collected and analyzed. The major findings obtained revealed that when tasked with a knowledge-production type of assignment, HEPG paid more attention to relevant information shown on screen and spent considerably less time on unrelated information, as compared to LEPG. This obtained result corroborated with the findings of McCrudden and Schraw’s study (2007), which indicated that proficient readers were cognizant of their informational needs and capable of zooming in on relevant information while detering from unrelated, unfruitful messages. It was also in alignment with the finding of Rouet and Britt’s research (2011) showing that proficient readers could target the most essential and pertinent information when given multiple pieces of text for better goal attainment (e.g., reading for understanding, task completion, etc.).

In addition, in this study, substantially more percentages of HEPG were found to satisfactorily
produce SGQ. This finding seemed to echo the findings of Yildiz and Çetinkaya’s study (2017) in that positive relationships exist between proficient readers’ attention and learning outcomes (e.g., comprehension, task completion).

4.1 Suggestions for Instructional Implementation

This study confirmed the findings of previous research that students at different English proficiency levels processed reading materials differently (in specific, attention allocation and management, as evidenced by fixations duration and visit counts) to better help meaningful input from the online reading material be noticed and become intake for learning (Schmidt, 1990, 2001). By purposively fixating on relevant, useful information presented on screen, successful task completion was ensured (as evidenced by SGQs satisfactorily meeting the criteria in this study).

On the basis of the current findings and in reference to existing literature on language learning, two suggestions are provided for instruction. First, instructors are advised to provide explicit attention cues and direction so as to effectively guide the attention of English learners with less language proficiency for focused, effective online reading and task completion. Second, instructors should simplify instructional screen message design presented to learners by taking out information not directly related to the assigned task for ensuring better attention management of low language proficient learners.

4.2 Limitations of This Study and Suggestions for Future Studies

Despite the promising findings of this study, the generalizability of the results should be considered limited due to its small sample size, short reading passages, and the difficulty level of the reading material. Future studies with a larger sample size, longer reading passages, and reading materials with different difficulty levels are needed. Also, mixed methods research by including other data collection methods, for instance, in-depth interviews or think-aloud protocols, can be considered for future studies. By so doing, insights on the reasons for attention allocation and management especially of the participants in LEPG regarding some areas of inquiry (e.g., why skipping viewing on-screen related hints, why fixating on unrelated hints presented on-screen, etc.) can be understood.

References


Exploring English Language Learners’ Conceptions of and Engagement in a Virtual Reality Learning Environment

Huilin WANG, Hanyong LIU, Zhihong LU & Chunping ZHENG*
Beijing University of Posts and Telecommunications, China
*zhengchunping@bupt.edu.cn

Abstract: With the progress of science and technology, virtual reality (VR) technology has attracted increasing attention in the field of education technology and language teaching. This inquiry explored English learners’ conceptions of learning and learner engagement in a VR learning environment. Participants were English learners from a comprehensive university in northern China. Data were collected from questionnaires and interviews. The results indicated that learners’ conceptions of learning English include eight factors, namely presence, motivation, extending, attention, interaction, understanding, obstructing conventional learning and diminishing imagination and learner engagement consists of four factors, including cognitive, behavioral, emotional and social engagement. Moreover, the present study found that there were several benefits and challenges of learning in a virtual reality environment.

Keywords: Virtual reality, conceptions of learning English, learner engagement

1. Introduction

The development of emerging technologies has increasingly influenced language education, which has been accompanied by a major expansion of research on Computer Assisted Language Learning (CALL) (Choi & Baek, 2011). Virtual reality (VR) technology has become a hot topic in second language studies by virtue of permitting researchers to design and construct the contexts according to educational objectives (Oblinger & Oblinger, 2005). What’s more, conceptions of learning and learner engagement play an important role in learning process and outcomes. Previous studies showed that conceptions of learning and learner engagement exerted a profound influence on learning process and academic achievement (Cheng, 2018). Based on the above research background, this study, from the perspective of individual characteristics of language learners, focused on college students’ conceptions of learning English and learner engagement in a VR learning environment.

2. Literature Review

2.1 Virtual Reality Technology and language education

VR technology was regarded as a powerful and promising tool in education because it could be pedagogically exploited through its unique technological characteristics (Mikropoulos & Bellou, 2011). Previous studies showed that learners improved their interest and learned effectively by learning target language in VR learning environments (Peterson, 2010). The design and construction of VR learning environments based on VR technology can provide learners with a realistic and situational language learning environment, promote learners’ communication and interaction, and enable them to attain an immersive learning experience (Hsu, 2017).

2.2 Conceptions of Learning and Learner Engagement in VR Learning Environment
The dimensions and features of conceptions of learning have been explored in early studies (Säljö, 1979). Recently, Cheng (2018) explored students’ conceptions of learning science by augmented reality (AR), including increasing presence, drawing attention, fostering motivation, extending content, attaining in-depth understanding, enhancing interaction, obstructing reading, and diminishing imagination. As an educational technology similar to AR, VR still lacks relevant studies on conceptions of learning. Fredricks and his fellows firstly defined that learner engagement included three dimensions: behavioral, emotional and cognitive engagement (Fredricks et al, 2004; Wang et al, 2011). In addition to the three components of engagement, Fredricks et al. (2016) added a social engagement dimension.

In order to explore the above issues, this research aimed at answering the following three questions: (1) What are the main factors of conceptions of learning English among college students in the VR-supported environment? (2) What are the main factors of learner engagement among college students in the VR-supported environment? (3) What are the benefits and challenges of the VR learning environment for English language learning?

3. Methodology

3.1 Research Context

The present study was conducted in a comprehensive university with emphasis on Computer Science and Technology, providing good technology support. The research took the course named ‘Learning English Through Culture: Viewing, Listening, Speaking’ as the research situation, combined the self-developed learning platform called ‘Situational English in Virtual Reality’ (SE in VR), and simulated the international airport as one of the main interactive contexts for situated English language learning. ‘SE in VR’ belongs to desktop VR where users can interact with the virtual reality environment using keyboard and mouse. It simulated the international airport for situational English language learning. There were ten learning tasks, encompassing word recognition, word matching, spot dictation, contextual interaction, rearranging order, short answers, compound dictation, summary writing, verbal report and role play.

3.2 Participants

A total of 69 EFL learners of this course participated in the study, including 52 males and 17 females, with an average age of 19. They came to the course on Tuesday or Thursday every week. All participants were non-English majors from different departments of this university. The 69 EFL learners used the platform to learn English, and then completed the questionnaires. Four of the participants were interviewed for the following qualitative research.

3.3 Data Collection and Analysis

This study was conducted in a mixed-method approach including quantitative and qualitative parts. A total of 69 participants in the study volunteered to respond to the questionnaires online, and all of them completed the questionnaires anonymously. 3. In addition, four participants were interviewed. Data analyses involved the following three phases. First of all, the construct reliability of the two questionnaires was conducted and the related descriptive data were reported. Then, learners’ responses to open-ended questions of the questionnaires were analyzed. Finally, the interview data were analyzed.

4. Results

4.1 Quantitative Results

4.1.1 The reliability analysis of CLEVE survey
Corresponding to the previous studies (Cheng, 2018), the results showed that there were seven factors generated with a total of 30 items. Table 1 demonstrates the mean and standard deviations for each item of the CLEVE survey, as well as the details of these items. The reliability coefficients for these factors were 0.97 (Presence, P), 0.95 (Attention, A), 0.94 (Motivation, M), 0.95 (Extending, E), 0.93 (Understanding, U), 0.92 (Interaction, I), 0.87 (Obstructing conventional learning, OCL), and 0.96 (Diminishing imagination, DI). The alpha coefficient of this study was around 0.87-0.97. Accordingly, these factors were considered to be sufficiently reliable to assess the students’ conceptions of learning English in the VR-supported environment. It should be remarkable that the factors of the CLEVE survey were divided into the two categories of positive and negative conceptions. Table 1 demonstrates that the students’ rating scores on the factors of positive conceptions are all higher than 3.7 points, indicating that the students of this study generally exhibited positive attitudes toward the VR-supported environment. Among their positive conceptions, they showed stronger conceptions of learning regarding increasing presence that learners experienced in the ‘SE in VR’.

4.1.2 The reliability analysis of LEVE survey

The results showed that a total of 35 items were presented and further grouped into four factors in Table 2. There were the mean and standard deviations for each item of the LEVE survey, as well as the details of these items. The four factors were ‘Cognitive engagement (CE)’ (α = 0.94), ‘Behavioral engagement (BE)’ (α = 0.96), ‘Emotional engagement (EE)’ (α = 0.95) and ‘Social engagement (SE)’ (α = 0.98). The alpha coefficient of this study was around 0.94-0.98, indicating satisfactory internal consistency of assessing learner engagement in the VR-supported environment. According to Table 2, the students’ rating scores on the factors of learner engagement were all higher than 3.2 points, indicating that learners had good engagement in the SE in VR. They were more engaged socially while they reported rating scores on the factors of learner engagement were all higher than 3.2 points, indicating that the students of this study generally exhibited positive attitudes toward the VR-supported environment. Among their positive conceptions, they showed stronger conceptions of learning regarding increasing presence that learners experienced in the ‘SE in VR’.

### Table 1. The Descriptive Data and Reliability Analysis of the CLEVE (n=69).

<table>
<thead>
<tr>
<th>Factors, specific items and descriptive data</th>
<th>Mean</th>
<th>S D</th>
<th>Factors, specific items and descriptive data</th>
<th>Mean</th>
<th>S D</th>
</tr>
</thead>
<tbody>
<tr>
<td>I think learning English in the SE in VR is to</td>
<td>3.90</td>
<td>.84</td>
<td>U-16 understand English knowledge in more depth.</td>
<td>3.93</td>
<td>.83</td>
</tr>
<tr>
<td>P-1 engage students in English language learning and apply it into the real scenes.</td>
<td>3.78</td>
<td>.94</td>
<td>U-17 increase understanding of language learning.</td>
<td>3.99</td>
<td>.76</td>
</tr>
<tr>
<td>P-2 help students learn and apply English from 3D perspective.</td>
<td>3.87</td>
<td>.84</td>
<td>U-18 deepen the understanding of cross-cultural communication.</td>
<td>3.79</td>
<td>.79</td>
</tr>
<tr>
<td>P-3 increase the feeling of being in real scenes.</td>
<td>3.96</td>
<td>.88</td>
<td>Factor 6: Interaction: Cronbach’s Alpha =.92</td>
<td>3.78</td>
<td>.80</td>
</tr>
<tr>
<td>Factor 1: Presence: Cronbach’s Alpha =.97</td>
<td>3.88</td>
<td>.83</td>
<td>I-20 create the opportunity for students to explore English language learning.</td>
<td>3.84</td>
<td>.82</td>
</tr>
<tr>
<td>A-4 attract attention.</td>
<td>3.86</td>
<td>.88</td>
<td>I-21 provide more opportunities to learn actively.</td>
<td>3.83</td>
<td>.77</td>
</tr>
<tr>
<td>A-5 enhance attraction of language learning.</td>
<td>3.83</td>
<td>.92</td>
<td>I-22 provide more opportunities to interact with the situational content.</td>
<td>3.87</td>
<td>.86</td>
</tr>
<tr>
<td>A-6 strengthen the students’ focus on the materials of learning English.</td>
<td>3.78</td>
<td>.94</td>
<td>Factor 7: Obstructing Conventional Learning (OCL): Cronbach’s Alpha =.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A-7 stop students from being distracted when they learn English.</td>
<td>3.80</td>
<td>.85</td>
<td>OCL-23 break the continuity of learning English.</td>
<td>3.20</td>
<td>1.09</td>
</tr>
<tr>
<td>Factor 3: Motivation: Cronbach’s Alpha =.94</td>
<td>3.80</td>
<td>.88</td>
<td>OCL-24 reduce concentration in learning English.</td>
<td>3.12</td>
<td>1.08</td>
</tr>
<tr>
<td>M-8 strengthen motivation of learning English.</td>
<td>3.81</td>
<td>.80</td>
<td>OCL-25 make us more willing to communicate with computers.</td>
<td>3.57</td>
<td>1.05</td>
</tr>
<tr>
<td>M-9 foster interest in learning English.</td>
<td>3.80</td>
<td>.87</td>
<td>OCL-26 make us not used to learning and applying English in real environment.</td>
<td>3.03</td>
<td>1.08</td>
</tr>
<tr>
<td>M-10 arouse curiosity in learning English.</td>
<td>3.88</td>
<td>.81</td>
<td>Factor 8: Diminishing Imagination (DI): Cronbach’s Alpha =.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M-11 enhance interest in real scenes.</td>
<td>3.90</td>
<td>.83</td>
<td>DI-27 have a negative impact on our imagination.</td>
<td>3.03</td>
<td>1.20</td>
</tr>
<tr>
<td>Factor 4: Extending: Cronbach’s Alpha =.95</td>
<td>3.90</td>
<td>.80</td>
<td>DI-28 reduce our imagination of real interaction.</td>
<td>2.94</td>
<td>1.15</td>
</tr>
<tr>
<td>E-12 extend learning content.</td>
<td>3.84</td>
<td>.80</td>
<td>DI-29 prevent us from thinking freely.</td>
<td>2.96</td>
<td>1.14</td>
</tr>
<tr>
<td>E-13 replenish English knowledge.</td>
<td>3.91</td>
<td>.84</td>
<td>DI-30 limit our imagination.</td>
<td>2.93</td>
<td>1.20</td>
</tr>
<tr>
<td>E-14 enrich the learning materials.</td>
<td>3.91</td>
<td>.76</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-15 provide students with more materials.</td>
<td>3.91</td>
<td>.76</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 2. The Descriptive Data and Reliability Analysis of the LEVE (n=69).

<table>
<thead>
<tr>
<th>Factors, specific items and descriptive data</th>
<th>Mean</th>
<th>S D</th>
<th>Factors, specific items and descriptive data</th>
<th>Mean</th>
<th>S D</th>
</tr>
</thead>
<tbody>
<tr>
<td>When I learning English in the SE in VR, ...</td>
<td>3.59</td>
<td>1.08</td>
<td>Factor 3: Emotional Engagement (EE): Cronbach’s Alpha =.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factor 1: Cognitive Engagement (CE): Cronbach’s Alpha =.94</td>
<td>3.59</td>
<td>1.08</td>
<td>EE-19 I’m looking forward to participating.</td>
<td>3.84</td>
<td>.90</td>
</tr>
<tr>
<td>CE-1 I will check the tasks carefully.</td>
<td>3.60</td>
<td>0.94</td>
<td>EE-20 I love to experience new technologies.</td>
<td>3.81</td>
<td>.96</td>
</tr>
<tr>
<td>CE-2 I will think about different methods.</td>
<td>3.65</td>
<td>.94</td>
<td>EE-21 I hope to know what I can learn.</td>
<td>3.90</td>
<td>.84</td>
</tr>
<tr>
<td>CE-3 I connect things to what I have learned before.</td>
<td>3.65</td>
<td>.94</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
did not match the software configuration requirements, resulting in system jam, crash and delay. At the same time, corresponding solutions were proposed. As for the system, the computer equipment since the SE in VR is still in its developmental stage, some problems encountered during the process.

### 4.2 Qualitative Results

#### 4.2.1 The benefits of the VR-supported environment for language learning

According to the subjective questions and answers in the questionnaire, it was found that 56 students had a good learning experience in the SE in VR. The results of sorting out their comments showed that their good learning experiences in the VR-supported environment were mainly divided into seven aspects. Table 3 shows all kinds of good experiences and examples of students’ comments.

<table>
<thead>
<tr>
<th>Description of learning experience</th>
<th>Examples of students’ comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning English in a new way</td>
<td>“It’s my first time to learn English in this way.”</td>
</tr>
<tr>
<td>Improving interests and motivation</td>
<td>“It improved my enthusiasm for learning English.”</td>
</tr>
<tr>
<td>Applying knowledge into practice</td>
<td>“I can put what I just learned into practice immediately.”</td>
</tr>
<tr>
<td>Having immersive experience</td>
<td>“The airport is so realistic that I immerse myself in it.”</td>
</tr>
<tr>
<td>Reducing the difficulty of learning</td>
<td>“It’s easier to learn English in simulated situations.”</td>
</tr>
<tr>
<td>Reflecting on their own shortcomings</td>
<td>“I found my lack of problem-solving ability.”</td>
</tr>
</tbody>
</table>

In a word, the study showed that the benefits of the VR-supported environment are not only created an authentic language learning environment, but also provided a new way which can promote motivation, apply knowledge into practice, enhance immersive experience, reduce the difficulty, and test shortcomings. All such benefits can help students to learn English more efficiently and effectively.

#### 4.2.2 The challenges of the VR-supported environment for language learning

Since the SE in VR is still in its developmental stage, some problems encountered during the process. At the same time, corresponding solutions was proposed. As for the system, the computer equipment did not match the software configuration requirements, resulting in system jam, crash and delay. Frequent crash caused the completed task schedule disappeared and needed to be restarted.

“Crashing three or four times caused a bad mood. When I finish eight or nine parts, the system may crash suddenly.” (Translated and excerpted from the focus interview transcripts of S4)

“The environment is too complicated. It’s a waste of time to find the designated items. I always get Chinese prompt after typing in the contextual interaction.” (Translated and excerpted from the focus interview transcripts of S3)

In short, there are still some problems in the VR-supported environment, including technical issues, unreasonable scene design, inconvenient operation and insufficient scaffolding. Students’
solutions are worthy of consideration. Their insightful suggestions are conducive to designing and developing VR-supported environments for EFL learners.

5. Discussion

5.1 Conceptions of Learning English in the VR-supported Environment

In this inquiry, a CLEVE survey was developed to investigate conceptions of learning English among college students in a VR-supported environment. The findings indicated that learners’ conceptions of learning English included eight factors. The instrument displayed similar factor structures as revealed by Cheng’s work (2018), and it showed satisfactory alpha reliability. The instrument validated in this research could assist instructors and researchers to gain an overall understanding of learners’ conceptions of learning English. According to the students’ rating scores in the survey, learners generally expressed positive conceptions of learning English while they might not have strong negative conceptions of learning English.

5.2 Learner Engagement in the VR-supported Environment

The present study investigated learner engagement in the VR-supported environment. Corresponding to the previous study of Wang et al. (2016), this study identified that learner engagement consisted of four factors, and showed the reliability of the LEVE survey. Consistent with the recent literature, the findings of this study supported learner engagement as a multidimensional construct (Wang & Degol, 2014). The results of this study demonstrated that learner engagement consisted of four theoretically distinct dimensions, and did not support recent research to regard learner engagement as a continuum (Sinatra, Hedy, & Lombardi, 2015). The multidimensional perspective of learner engagement provided a richer characterization of how learners behave, think, feel, and socialize with others in the SE in VR, rather than exploring each of the dimensions separately (Wang et al., 2016). In this study, learners generally had positive engagement in learning English in the VR-supported environment.

5.3 Benefits and Challenges in Applying VR to English language Learning

Plenty of literature supported the effective use of the VR-supported environment in foreign language learning because of its potential of providing learners with the suitable language contexts and the possibility of enhancing learners’ language competences (Lan et al., 2013; Peterson, 2011). The results from this study indicated that the beneficial characteristics of VR helped learners learn English in creating an authentic learning environment and enhancing immersive experience. Immersive simulations made the language contexts more realistic, resulting in heightened involvement and positive learning outcomes (Liou, 2012). Although students encountered many unexpected challenges during the learning process, most students are positive on learning English in the VR-supported environment because of their good learning experience. They hope that the SE in VR can be continuously improved and put into practical teaching.

6. Conclusion

This study investigated learners’ conceptions of learning English and learner engagement in a VR-supported environment. The findings indicated that learners’ conceptions of learning English in the VR-supported environment included eight factors, namely presence, motivation, extending, attention, interaction, understanding, obstructing conventional learning and diminishing imagination. Learners generally expressed positive conceptions of learning English. The relatively stronger factor of conceptions of learning English was presence, while they tended to consider learning English in the SE in VR as obstructing conventional learning. What’s more, learner engagement in the VR-supported environment consisted of cognitive, behavioral, emotional and social engagement. Among their engagements, learners reported more social participation in SE in VR. This study also explored benefits and challenges in applying VR to English language learning and provided suggestions to design and
develop VR learning environments in the future. However, the sample of this research was constrained, so future studies could consider a larger group of students with different background and this study only analyzed a part of the data with the limited time. Follow up studies concerning relationship between conceptions of learning and learner engagement are expected in the future.

Acknowledgements

This research is supported by the Teaching Reform Project of Beijing University of Posts and Telecommunications (2019Y003) and the Project of Discipline Innovation and Advancement (PODIA)-Foreign Language Education Studies at Beijing Foreign Studies University (2020SYLZDXM011).

References

Examining primary students’ after-class vocabulary behavioural learning patterns in user-generated learning context: a case study

Yanjie Songa*, Hiroaki Ogatab, Yin Yanga and Kousuke Mouri

aThe Education University of Hong Kong, Hong Kong
bKyoto University, Kyoto, Japan
ysong@eduhk.hk

Abstract: In this paper, we explored the primary students’ after-class vocabulary behavioural learning patterns using a mobile user-generated-content (m-UGC) tool. A total number of 21 Grade 4 students from an elementary school in Hong Kong were recruited. Case study approach was adopted. Data collection included students’ learning logs on the m-UGC tool. Content analysis and visualization were used to analyze data. Three sub-groups of primary students’ after-class vocabulary behavioural learning patterns were identified and future studies were explored.

Keywords: vocabulary learning, visualization, behavioral patterns

1. Introduction

Technology plays a vital and rapidly evolving role in the area of language learning. Vocabulary learning is one of the major challenges foreign language learners confront during a second language acquisition. Despite a growing body of research that examines the effect of technology on students’ vocabulary learning performance, some have reported the benefits of technology in students’ vocabulary learning (Huang, Huang, Huang, & Lin, 2012; Ali Mohsen, 2016; Franciosi, 2017). However, little is known on students’ vocabulary behavioral learning patterns in authentic learning context. Rarely explored are studies aiming at tracing students’ online vocabulary learning which captured real-life learning experiences in primary school education, which is the focus of this study.

2. Literature Review

2.1 English as a second language (ESL) vocabulary acquisition

There are varied approaches to ESL vocabulary learning. Studies on ESL vocabulary learning show that review, consolidation and application of words are critical for learners’ vocabulary retention (e.g., Ma, 2014). “Good” language learners usually are involved in more out-of-class learning activities to practice the target language (Reinders, 2014). However, the opportunities for Hong Kong school students to learn English outside of classroom are rare (Chik, 2015).

2.2 Technology-enhanced ESL vocabulary acquisition in real life learning environments

In the digital age, studies on using technology, especially mobile technology and ubiquitous learning systems to enhance vocabulary learning in real life learning environments beyond the classroom are on the rise. Some studies have researched on investigating learners’ vocabulary learning using artifact creation approach to allow learners to create meaningful pictures that are associated with the new word to enhance their vocabulary retention (e.g. Foomani & Hedayati, 2016). Adopting the approach, learners are provided with opportunities to choose the object they prefer, take a picture and link it to the new word learned in class in real life. In addition, in some studies, learners are also allowed to describe the context of taking the pictures related to the word and comment on and share their created artefacts (e.g., Mouri Uosaki & Ogata, 2018). Such learning process involves learners’ choices for autonomous learning (Godwin-Jones, 2019). However, few studies have examined learners’ learning strategies by
understanding their behavioural learning patterns through learning analytics at a primary level, which can inform pedagogical decision refinement for future pedagogical practices.

2.3 Learning analytics and language learning

Godwin-Jones (2017) posits that data has played a significant role in the field of second language acquisition. With learning analytics, educators are able to be better informed with timely decision-making in refining the instructional design for improving pedagogical practices. For example, Hsiao, Lan, Kao, and Li (2017) developed a visualization analytic approach to understanding the impact of various learning strategies on college students’ Chinese vocabulary in a virtual world. The research findings show that the visualization analytics method could help teachers visualise students’ different learning strategies of vocabulary acquisition. Some studies have shown the effectiveness of learning analytics in a computer-assisted vocabulary learning context. For example, Mouri, et al. (2018) reported a study on evaluating the effectiveness of a learning analytics tool in connecting students’ vocabulary learning acquired via eBook to those learned from real-life in higher education and the results were positive. Although learning analytics has been increasingly used in language learning especially in vocabulary learning, practical implications from extracting learners’ behavioural learning patterns to inform pedagogical decision making in informal learning environment have rarely been explored. In addition, a few studies on adopting learning analytics to examine learners’ behavioural patterns in language learning have been conducted in higher education.

In light of these issues, this study aimed at examining learners’ vocabulary behavioural patterns out of class on a ubiquitous learning system via learning analytics at a primary level. The research question is: What were the students’ after-class vocabulary behavioural learning patterns on m-UGC tool?

3. This Study

3.1 Participants and Learning Topics

A one-year case study approach was adopted in order to inquiry into the phenomenon of students’ vocabulary learning process in real life context and identify their learning patterns (Yin, 2002). A sample of 23 students (4C01- 4C23) were recruited from one class of fourth graders at an elementary school in Hong Kong. The mean age of all the subjects is 9.5 years old. The study lasted for two weeks. English was used as the medium of instruction. To consider the research ethics of study that involved collecting data from the participants, a written informed consent form was obtained from both the participants and their parents,. Two students withdrew the study. In the end, 21 students were included in this case study. The topic of vocabulary learning reported in this paper was “places”.

3.2 The m-UGC tool

The m-UGC tool is adapted from the System for Capturing and Reminding of Learning Log (see Ogata et al., 2011; Song & Yang, 2019) was used in this study. Students could add learning notes, upload pictures, describe the newly acquired vocabulary and comment on peers’ learning logs by giving text-based comments on SCROLL (refer to examples in Figure 3 in results and discussion section).

3.3 Coding Scheme of students’ behavioural learning patterns

Based on the main behaviour categories in terms of students’ posting, descriptions and social behaviors (comment peers’ postings on the m-UGC tool), three subgroups were identified: active social, active non-social and passive users by referring to the literature (Gerson, Plagnol, & Corr, 2017). Students who were active in the creation of comments were categorized into “social” users, and active users were those in the high engagement with the creation of learning logs (posting pictures and vocabulary) and descriptions.
Firstly, seven students were categorized as “passive users” because they did not post any comments and in which two students only post one posting respectively. The rest 14 students were divided into “active social” and “active non-social”. Secondly, given the fact that 125 picture-and-text based postings, 90 text-based descriptions and 26 comments were collected in this study. The criterial of dividing “active social” and “active non-social” users was: if the number of students’ postings were larger than 8, he/she could get 1 mark, otherwise they would get 0; in addition, students could get 1 mark if they made more than 6 pieces of text-based description, otherwise they would get 0; besides, students could get another 1 mark if they commented on others’ learning logs. Hence, students’ posting score ranged from 0 to 3 marks. Students who got 3 marks were categorized into the “active social” group. The rest of students were “active non-social” users. Finally, each subgroup consisted of 7 students.

Data Collection and Analysis

Data sources included students’ vocabulary learning logs recorded (pictures, descriptions and comments) on the m-UGC tool. Both qualitative and quantitative data analysis methods were adopted. Firstly, student created vocabulary learning logs were counted, coded and categorized into postings (picture and vocabulary), descriptions and comments. Secondly, students’ vocabulary learning logs were analysed using Gephi (https://gephi.org) to identify students’ behavioural learning patterns.

Results and Discussion

The results of the students’ after-class behavioural learning patterns on the m-UGC tool were presented in this section. There were 125 vocabulary learning logs (pictures and vocabulary) in total created by the students with 26 comments documented on the m-UGC tool. Among the 125 postings, 118 postings (94.4%) were related to the vocabulary learned on the topic of “Places”. The rest of 7 postings were not included in the vocabulary list. The distribution of students’ behaviour categories among three sub-groups are listed in the Table 1.

Table 1. The distribution of students’ behaviour categories across three sub-groups

<table>
<thead>
<tr>
<th></th>
<th>Posting</th>
<th>Description</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active social</td>
<td>72 (57.6%)</td>
<td>70 (77.8%)</td>
<td>24 (92.3%)</td>
</tr>
<tr>
<td>Active non-social</td>
<td>51 (40.8%)</td>
<td>19 (21.1%)</td>
<td>2 (7.7%)</td>
</tr>
<tr>
<td>Passive users</td>
<td>2 (1.6%)</td>
<td>1 (1.1%)</td>
<td>0</td>
</tr>
</tbody>
</table>

Students in sub-groups of “active social” and “active non-social” were active in creating picture-and text-based learning logs, while 77.8% of text-based descriptions were made by active social users, 21.1% were made by active non-social users. Females were dominated in the “active social” group, accounting for 71.4%, in contrast, males were more likely to be “passive users” (see Table 2).

Table 2. Gender distribution across three sub-groups

<table>
<thead>
<tr>
<th></th>
<th>Active social</th>
<th>Active non-social</th>
<th>Passive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Percent</td>
<td>Frequency</td>
</tr>
<tr>
<td>Female</td>
<td>5</td>
<td>71.4%</td>
<td>4</td>
</tr>
<tr>
<td>Male</td>
<td>2</td>
<td>28.6%</td>
<td>3</td>
</tr>
</tbody>
</table>

Gephi (https://gephi.org) was used to analyze students’ learning logs and peers’ connections on the m-UGC tool. The visualization was done using Gephi’s default layout algorithm (Fruchterman reingold), which renders each node’s position according to its relations and connections (Jacomy, Venturini, Heymann, & Bastian, 2014). First, students’ logged data on the m-UGC tool was transformed into data that could be read by the visualization tool Gephi. As showed in Figure 1, 175 nodes were created, in which 21 nodes represented 21 participants, and 152 nodes of target words. Gephi created each edge to connect students and learning logs. For example, if 4C01 learner created word “Library”, node 4C01 was connected to the node “Library”. Figure 1 shows the relations between
each student and each word that they were logged on the m-UGC tool. The size of node represented the “degree centrality” (its number of connections). Hence, 4C02, 4C01, 4C09, and 4C06 were in higher degree centrality than other students, indicating that these students were active in contributing their learning logs.

Figure 1. Visualization of students and learning logs

Take student 4C02 as an example (see Figure 2). The value of degree centrality was 19 as calculated by Gephi 0.9.2, meaning student 4C02 created 19 words and was the most active in posting learning logs in the class. The value of degree centrality indicated the number of edges connected to a node, with higher degree values relating to more influential nodes (Hernández-García, González-González, Jiménez-Zarco, & Chaparro-Peláez, 2016).

Figure 2. Visualization of student 4C02’s learning log

The most active learners 4C02, 4C06, 4C09, and 4C01 contributed 44.8% of the entire class’s postings and created 56 learning logs in total. We also found that the learner 4C02 posted words related “restaurant” that other student did not post, such as “Subway”, “Subway box”, “Subway cookies” and “Subway menu roblox”. It was worth mentioning that “Subway” is a fast food restaurant and is common in Hong Kong. Therefore, the learner 4C02 could relate pictures and vocabulary in real life context to what he/she learned in class (See Figure 3).

Figure 3. Examples of Student 4C02’s learning logs related to “restaurant”
As for posted vocabulary, words such as “library”, and “police station” were in centrality among 21 students, indicating that 11 of 21 students recorded word “library” and 9 students recorded the phrase “police station” (refer to Figure 4).

Figure 4. Example: visualization of words “library” and “police station”

The visualization analytics also effectively displayed the peers’ connections on the m-UGC tool (Figure 6). The directed relationships were represented by arrows pointing from a source node to a target node. For example, if 4C10 commented 4C01’s learning log, the source node would be 4C10 and the target node would be 4C01. The thickness of lines implied the average number of peers with which each learner interacted. Active social users created much thicker lines, indicating that two contributors were more likely to interact with each other as showed between student 4C01 and 4C10. Figure 5 provided information about which students were not involved in commenting on the m-UGC tool represented by unlinked nodes.

Seven students (4C01, 4C03, 4C07, 4C09, 4C10, 4C15, 4C17) were active in posting 26 comments in total. The size of 4C01’s node was the biggest in the class, indicating that 4C01 received the most comments from peers, in which the interaction between peers with 4C10 was in the highest frequency. In addition, as showed by the arrows, 4C01 had interactions with 4C03, 4C06, 4C09, 4C10, 4C13, 4C14, 4C15 and 4C17.

Figure 5. Visualization of peers’ interactions on the m-UGC tool

5. Conclusion and Future Studies

This study aimed at investigating students’ after-class vocabulary behavioural learning patterns on a user-generated-content (m-UGC) tool by tracking their learning process through learning analytics and visualization.
Based on students’ number of posting, description and comments, three sub-groups were identified: active social, active non-social and passive users, which can help teachers make pedagogical-decision-making in future pedagogical practices. This study had limitations. Considering the tentative nature of this study, we only reported 21 students’ learning logs under one topic “places”. In addition, the participation level in commenting of students was not satisfying, we should identify the reasons and collect more comments in the future studies. In the future, we will increase the sample size, explore the impact of the patterns on students’ vocabulary learning performance and the relationship between students’ after-class behavioural learning patterns and students’ vocabulary performance will also be investigated.

Acknowledgements

The study was funded by the General Research Fund (Ref. 18611019) by Research Grants Council, University Grant Committee, Hong Kong.

References


Learning on Country: A Game-Based approach towards preserving an Australian Aboriginal Language.

Cat KUTAY*, Deborah SZAPIRO, Jaime GARCIA, William RAFFE

*Top End Language Lab, Charles Darwin University, Australia
bUniversity of Technology Sydney, Australia
*cat.kutay@cdu.edu.au

Abstract: Nginya naaa-da bangar-mari dalang-wingaru-dane. Ngyina diya-ma murru dalan-wa dalang-ra'. This paper presents the design of a prototype 360 degree, interactive, Indigenous language learning game to support the reclamation of Indigenous languages through immersion in community oral traditions expressed through visual and audio effects and the choreography of the characters within the game. The project is underpinned by a foundational acknowledgement that Aboriginal culture is held within the country specific to the language and embodied in that country’s landscape. Learning within the game is based around themes of country, weather, local environment and kinship. Animation and design principles were applied from an embodied communication perspective to increase engagement and to reinforce language learning principles, with Indigenous animation and design students bringing an Indigenous perspective to the gestural and design content of the game.

Keywords: Game design, Design Process, Social design, Language in context

1. Introduction

You can have words written on a page all day but they don’t make any sense until people speak them. And once people pronounce...it lives, it breathes. Green (2018)

This paper presents a prototype game for the Learning on Country project, an interactive Indigenous learning game for the reclamation of languages. The project frames innovation as a holistic system that balances cultural, social and technological perspectives to provide embodied communication for learning. The language is Dharug of Sydney and the country depicted is the land which the university occupies. This was motivated by discussions with Dharug speaker and language teacher, Richard Green, who wished to hear his language spoken around the streets of Sydney and builds on previous work on using technology to support community knowledge sharing (Kutay & Green, 2013).

Languages are an essential part of the knowledge system of human culture and frame people’s way of seeing and understanding the world. As Nettle and Romaine (2000) have shown, the loss of diversity in languages and the knowledge and practices contained within a language has accelerated the loss of diversity in the environment where these languages developed. Lewis and Simons (2016) state that language use in the community, outside formal institutions, is critical for reclamation to be successful. The valuing of the knowledge that is spoken through a language keeps a language alive, and the ongoing practice of language in community reclaims the knowledge that lost since colonization.

2. Supporting community knowledge sharing spaces through the use of technology

The project was developed with community members to share information about their cultural values and environmental knowledge as a way of restoring their identity and traditional language use. Game-based learning has shown to be effective in teaching for modern languages (Klimova, 2015) and younger generations have grown up with game culture, bringing a shared code of meaning when approaching a game interface. Online oral recordings and text worksheets provided the material for

---

1 We look to make material to teach language to all. We do this in a way that all languages can talk this way.
creating a multimodal, immersive and interactive context for learning language, that incorporates community traditions through location on country and visual choreography of the characters.

When teaching Aboriginal languages there are some required components that can be achieved in games. Firstly, the language needs to be embedded in the culture (Furstenberg, 2010) as this understanding is necessary to engage with the language. Secondly, there is fun and enjoyment. The approaches of action-based learning and student agency in their own learning are finding a place for these languages (Fang, n.d.). Thirdly is a need for repetition for reinforcement. We use a set wordlist, keeping the changes to a minimal, with a restricted vocabulary and set grammar components per lesson (Sharpe, 2019). The Aboriginal language phrases are based around themes that are important for sharing cultural knowledge and are part of everyday conversations that can be repeated with other speakers.

The game is multi-platform, animated in both 2.5D and 3D and developed in Unity. A large screen 360-degree format allows for full immersion, while community access to the game on an ordinary PC is possible through a WebGL export. The game is designed as a developing conversation which supports players and is able to expand on themes in new levels to emulate cultural knowledge sharing practices.

3. Related Work

An important aspect in the design was future re-use of the system for different local languages. The lack of linguistic aspect to provide rules for reclaimed languages make any process of script generation difficult to verify for such languages, so the generation of the script was an important area in which to involve speakers (cf. Jaschek et al, 2019). Similarly, auto-generating levels (Gaisbauer et al 2019) or generating maps in terms of the geometry and the density of content within that geometry (Raffie et al, 2014) would remove the authenticity of the location to the Aboriginal traditional environment which is important in this work. This means the reusable aspects are the gaming aspects, how the player interacts with the landscape and the language, and how that language is communicated to the player. The increased use of animated characters in TV shows (Little J & Big Cuz) and machinima techniques to generate scenes for games from existing animated characters and scripts (Anjos et al 2013) may be useful in future with more resources developed in this area.

Language teaching on mobile apps, in stop-motion animations (McKnight et al, 2010) and in CMS learning systems (Bow, 2019) are all bringing language to life online. Apps usually provide wordlist and sometimes example phrases which are used to create exercises such as ‘select the meaning of a word’, ‘listen to this word and select the written form’ and ‘fill in the word’ (Yawuru, n.d.). Stop-motion animations have been popular with languages which have more speakers and resources in central and northern Australia to create a knowledge repository of the language. Language centres in Australia have started to provide online structures for teachers to develop material in class with their students, based around thematic issues that focus learning-limited word lists each day (Sharpe, 2019). Virtual reality in non-first language learning has often been successful due to its immersive nature and ability to authentically represent the culture of the language environment (Lan, 2020; Yeh and Lan, 2018).

This project expands on this work by using computing methods to enable gaming and repetition techniques in a template interface. The gaming system and interactions are generic to a language learning environment and can be applied to specific Aboriginal languages by inserting local imagery of country while using the spoken and written teaching examples available on websites like http://daraq.dalang.com.au. In this way, we are able to emulate the classroom learning strategy used in language centres, where only Aboriginal language is used, in a context that emulates aspects of everyday life for the students, placing them in a ‘virtual’ on-country context.

4. Methodology

As discussed, the project frames technological innovation as a holistic system which works with a balance of cultural, social and technological elements in equal consideration. It uses an Indigenous methodology that ensures research pertaining to Indigenous issues is always underpinned by an Indigenous perspective (Smith, 1999). The project employed three Indigenous student animators in developing the resources to express the language and knowledge in an authentic way. It is an integral
aspect of the methodology and one that should be considered to ensure engagement and agency in the co-creation of all cultural games and in particular in language teaching material by Indigenous community members as it maintains culture and supports reduced reliance on external developers.

Community speakers produced phrases that would be in common use in an oral language community. In keeping with Richard Green’s philosophy of learning phrases, not just words, the game avoids starting with wordlists and building up to phrases. This encourages the person learning to understand the words in context. The player who is starting their learning can pick out common words that reoccur in relation to different aspects of the gaming environment.

Gamification is based on navigating the landscape to hear the language and select places in the virtual world, as well as the ability to select the mode of learning, i.e. immersion; language with Aboriginal subtitles; and language with English subtitles. Features added to help convey the context and topic of the conversation such as short animations for each audio segment provide gestural cues for what is being said. Using the interactivity of the game, players can repeat phrases or select to move on.

A 2.5D environment was designed to situate the language in its specific country of origin. The current environment is based on the Sydney region as the language we are working with is of this area, although the environment can be re-used for coastal languages ranging from Newcastle to Wollongong. This environment includes land, water, plants and animals and animation of significant features.

5. Design and Animation

The aesthetics are designed to bring an Indigenous perspective to language learning. As country is important to Indigenous languages, we looked to create an environment that spoke to the area of the language, which was Sydney. This is shown as a Google extract in the first scene. However, to be on the land and talking about it was difficult as there were no images for pre-colonised Sydney.

The background for the localised scene was constructed using the work of painter Joseph Lycett who was one of the first artists to paint the Australian landscape without overlaying an English scene. He is notable for his documentation of Aboriginal people in his work, which forms some of the first figurative observation of Aboriginal culture. This both anchored the game in country and also brought a shared code of meaning to the game as anyone from the Sydney area would recognise and relate to aspects of the landscape yet also see what has been lost with European development.

Animation, choreography and imagery was designed from an embodied communication perspective so that participants can understand the meaning of phrases from the body language of the speaker. Three Indigenous UTS Bachelor of Design in Animation students were employed to design characters (Joshua Yasserie), 2D animation (Genevieve Stewart and Jalama Towney) and 3D elements (Jalama Towney). Stewart directed the majority of the gestural content of the characters to bring a deeper level of understanding of the language phrases through the characters’ body movements.

5.1 Language Themes

The phrases were recorded by Dharug speakers who have some proficiency in the language. The script of the game was designed to practice specific features of the language under themes. Sentences were structured for gradual development of language around the themes of each location. The scenes are:

1. **Welcome**: A welcome to country is highly personalised but has some standard components. We use this to open the game and credit the makers of the language resources and animations.
2. **Country**: We use an overview of the landscape to talk about country. This scene discusses language for naming country; simple verbs for movement; pronouns and locatives
3. **Weather**: As a common topic of conversation a short prototype on cloudy weather was included. This introduces some nouns, adjectives and continuous tense.
4. **Local environment**: When the birds lands the words are of the immediate world around us, part of which is a bora ground or place to educate community about caring for country. This scene is to develop directions and imperatives.
5. **Kinship**: In an introduction to any language or community we put the player in a relationship to the others in the game. This introduces names of family and asking questions
5.2 User experience

The animations and choice of interactions aim to reduce dependency on text. The first time the game is played through there are no subtitles with the audio. As most speakers will not know the language, on repeated play we provide Dharug then English subtitles.

![Image](Ngaya_wumara-we_Gadigal_nara)

*Figure 1* Magpie flies over Gadigal land towards Wulumulu - place of whales

The magpie was chosen as a Dharug totem to open with talking about the country (see Fig. 1). The next scene focuses on the clouds and the area surrounding the magpie. This scene could be extended into the night, where the stars can be shown using Stellarium software, from the Sydney Observatory, to present local stories about the stars. The player stays as the magpie which moves from the 3D World to the 2.5D scene and lands. Each of the animals and human characters speak about themselves and to each other (see Fig. 2). Background sounds set the scene and draw attention to areas of the landscape.

6. Learning approach

The gamification elements provide a pleasant learning environment including the ability of the player to fly around the landscape as the bird, talking of its country. The player selects to move to the next phrase and when they come to discussion of the weather, animals in the scene or locations in the scene, these features are enacted or highlighted in the game. When the magpie passes over important locations in the Sydney landscape this is described in language. At the university, a site of an old Bora or ceremonial ground, the magpie enters the next scene. The player is able to look around the scene and follow actions by avatars and their attention is directed to aspects of the scene by sounds.

![Image](Magpie_introduces_Kangaroo_by_name)

*Figure 2* Magpie introduces Kangaroo by name

A cultural aspect that is included is the use of a story or short narrative at the start to position the listener in the environment. This includes some of the words of the lessons so that as they go through the game, “information” provided at the start can become “knowledge” when experienced in a simpler context (Verran, 2007). Additional resources on the language such as the original material in lesson format can
be found on the associated language website. In bringing these resources together with an immersive learning environment, we provide a format for repetition of the website worksheets, allowing the online teachers to generate the phrase sequences to be learnt, which can be represented in the game. In future other languages can use the web system to provide phrases in a new language through online worksheets on their own version of the website, with the Application Programming Interface (API) to provide in the game. We envisage that other languages would want to include discussion of other features of the landscape which the speakers are familiar with. For this work new animations would be required.

7. Discussion

7.1 Feedback on authentic learning

As part of the ongoing design of the interface, one participatory design workshop was run with the full 360 surroundings before the COVID-19 shut down. This involved people interested in teaching and learning Aboriginal languages, including a language speaker. Due to their advanced knowledge of the language the speakers were not able to judge how a novice would respond, but they appreciated the use of language subtitles in the game as a learning aid. The use of animations to emphasise what was spoken, and the realistic landscape presented were considered an advance in language games for the participants, although they did not indicate any explicit effect that helped communicate the sense of the language spoken, other than the entire environment. This required further research.

We exported a web-based version of the game for people to review and this will be available for display at the conference. The user testing of players has come from this version, which is less immersive but still an embedded format of the knowledge and relies on a Chrome browser. We ran a survey on the user experience of the animations, gestures, sounds and subtitles, and how they reinforced the language. We received feedback from 10 participants from linguists to young language learners.

The game is designed to reinforce lessons, so the experience was a steep learning curve. Familiar words were noted as setting the context of homecoming and location (3d scene) and greeting (2.5D scene). The beauty and immersive quality of the environment and the background sounds was found engaging, and some noted the gestures brought the language to life. Participants commented that the interplay of all the components created a sense of being there. This was a single session and it would take continued repetition and experience to gain language competency. Key feedback in both formats noted the value in seeing the language brought alive. Working with the language learners we saw a shift from the need to express realistic conversation content to one of supporting learners via a slower and more repetitive language progression. This could be strengthened by extending and adding scenes.

7.2 Expanding the system

The oral resources for the Dharug language used in the game were collected by a Dharug speaker as lessons on the website. Developing language teaching material is time consuming. The ability to re-use this template lends itself to reinforcement learning in other languages. The API that is present on most web systems allows us to re-use and reapply language material from a website into apps and games to improve learning. The gaming format reduces reliance on text-based formats for learning language in what are traditionally active and oral learning cultures. Much of the work has been to develop the interaction between the object in the game, which can be added to any new objects for extension of this game or for new languages. The process has provided a model for integration of future audio scripts and the components needed to represent the new contexts:

- **Background**: New scenes can be inserted into the game. This will include integration of Stellarium into the game and addition of different skyboxes;
- **Dialogue**: Will be spoken by characters with expression to match the text;
- **Animations**: The present animations are human avatars and representations of animals common to many areas. These animated resources can be expanded to accommodate additional dialogue;
- **Avatar**: Introducing an avatar for the player will allow responses to trigger from characters;
- **Sound**: Different language options may be linked to a character in the same setting.
8. Conclusion

This paper has discussed an innovative design approach that was achieved through a holistic balance of cultural, social and technological systems. It outlines the process of designing a prototype interactive Indigenous language learning game aimed at the reclamation of Indigenous languages, and a process that supports cultural and community knowledge sharing spaces through the use of technology. The holistic design and integration of each element of the game was a significant factor in the success of the project for immersion and cultural learning contexts. The game links with work produced by language teachers through adapting lessons to this new environment, in order to reinforce learning and provide a way to bring the language back to its country.

References


TELL for Indigenous Australian languages

Catherine BOWa,b*

a Charles Darwin University, Australia
b Australian National University, Australia
*cathy.bow@cdu.edu.au

Abstract: The area of technology-enhanced language learning (TELL) has traditionally focused on languages with many speakers and multiple resources available. However, there are numerous affordances of TELL which can support smaller language groups to prepare and deliver language learning programs. This paper discusses some of the issues relating to TELL in the context of Indigenous Australian languages, particularly how Indigenous communities can use TELL to support language work. It goes on to describe a project designed to create an online template using low-cost and low-tech tools. While developed for a specific context – creating online courses for teaching Indigenous languages at university – the Digital Language Shell has a range of possible applications to maintain Indigenous languages under Indigenous authority.

Keywords: Australian Indigenous languages; online learning; Digital Language Shell, computer-assisted language learning

1. Introduction

The area of technology-enhanced language learning (TELL) has traditionally focused on languages with millions of speakers and multiple resources available. However, there are numerous affordances of TELL which can support smaller language groups keen to share their language and culture online. This paper discusses some of the issues relating to TELL in the context of Indigenous Australian languages, particularly how communities can use TELL to support their own aspirations for language work. The variety of language ecologies in Australia means that there is no single tool to meet all needs, however particular questions can be asked to determine appropriate strategies.

Existing TELL tools may not be suitable for Indigenous language contexts, and development of new tools generally requires significant financial and technical investment. This paper describes a project designed to consider the possibilities of creating an online template using low-cost and low-tech tools that Indigenous communities can use for free and with minimal technical proficiency. Consideration of a number of issues – audience, cost, language resources, archiving, access, etc – influenced the development of an online template, now available as the Digital Language Shell (https://language-shell.cdu.edu.au/). While designed to support teaching and learning of Indigenous languages, such a tool can also innovate ways of sharing Indigenous knowledge practices, record endangered language practices, document different dialects or varieties of a language, and incorporate cultural knowledge.

The paper begins with an overview of the Indigenous Australian language context, then outlines the benefits of TELL, and issues relating to development of TELL resources in this space, then goes on to describe the Digital Language Shell and its users. The paper provides insight into an area that is little explored in the literature, with regard to supporting the teaching and learning of less widely taught languages through TELL.

2. Benefits of TELL in the Indigenous Australian language context

While it is estimated that between 250 and 750 distinct languages were spoken in Australia prior to colonisation, the most recent survey (National Indigenous Languages Report, 2020) states that only 123 are still in use at some level. Of these, only 12 are still considered ‘strong,’ meaning they are still spoken by all age groups and being passed on to children.
Around 3.3% of Australian population identify as Indigenous (approximately 650,000 people), of which only 10% report speaking an Indigenous language at home (Australian Bureau of Statistics, 2016). These are largely traditional languages, but also include the increasing use of new or ‘contact’ languages such as Kriol or Yumplatok. Australian language ecologies are constantly in flux, due to the natural process of language change, the impact of policy on language practices in Indigenous communities, and the role of English in Australia.

Technology is commonly used for language documentation and description, which can support the work of revitalisation of endangered languages (Bow, forthcoming; Galla, 2016), however the use of technology to support teaching and learning Indigenous languages are less evident. Some of the recognised benefits of TELL programs, such as improving student multimedia learning experience, enhancing learner autonomy and widening participation (Yang & Rau, 2005), are highly relevant for the Indigenous context. In particular, TELL can transcend geographic distance and enable alternative pedagogical approaches. Australian languages are intimately connected with land, so technology can enable language authorities to stay on their land and teach learners in other locations (Christie, 2010).

TELL generally relies on languages with large speaker bases and significant resources, however there are moves towards drawing on the affordances of computational linguistics and natural language processing (Nerbonne, 2005) to work with low resource languages (Ward, 2015, 2017). Such work requires collection of large corpora to enable processes of speech recognition, machine translation, etc., which are currently not available for any Australian language, however there are efforts in this area (Centre of Excellence for the Dynamics of Language, 2019) and applications becoming available (Foley et al., 2018; Lane & Bird, 2019).

3. Developing TELL for Indigenous languages

The wide range of contexts in which technology can be used to support Indigenous language teaching and learning means there is no single use-case which could address all the needs of different communities or educational contexts. Careful attention to issues such as audience, cost, availability of resources, and access are needed when considering options, as well as taking seriously the cultural authority, protocols and practices of the local Indigenous community of engagement (Christie, Guyula, Gurruwiwi, & Greatorex, 2013).

Potential audiences in both formal and informal contexts include school curricula and community programs, which may involve some heritage learners of the language or may have mostly non-Indigenous learners. University courses may focus on the needs of those working in the language area, either professionals (e.g. in health, education, justice) interacting with speakers of the language, or as researchers with interests in the region (Simpson, 2014).

With small audiences for TELL tools in this context, there is little commercial interest from popular language teaching platforms such as Duolingo and Rosetta Stone. High-quality learning management systems (LMS) such as Blackboard and Moodle may not be appropriate for the Indigenous language context (Hugo, 2014). Indigenous communities and groups vary in their capacity and resources, with some having access to large commercial or institutional platforms for developing and sharing resources, and others needing to seek less expensive options. Some existing apps and tools for language learning can be customised for Indigenous languages, however they may impose a particular academic style that is not compatible with Indigenous pedagogies, or may use graphics and language themes that make them inappropriate for this context (Galla, 2017; Holton, 2011).

Creating new tools from scratch can be expensive, however there are some free and low-cost options which are supported by a community of users, rather than requiring maintenance contracts at additional cost. There may be costs associated with online hosting and storage of resources, and costs saved on buying or maintaining existing tools are likely to be spent on time developing and customising the tool. The greatest cost is likely to be in human resources, so for Indigenous communities with limited resources, the main expense should be paying Indigenous people to develop materials and design curricula, to support local economies and skills development.

There is great variety of the quality and quantity of materials available for teaching and learning Indigenous Australian languages, from a few historical documents to rich grammatical descriptions (Gaby, 2008), and dictionaries (Goddard & Thieberger, 1997). Other sources of material, such as vernacular literacy materials (Bow, Christie, & Devlin, 2017; ‘Living Archive of Aboriginal
Languages’, 2012), locally produced videos, blog posts, signage, children’s stories, musical recordings and other multimedia content, can be identified and examined for content to explain or demonstrate particular linguistic or cultural concepts. New resources can be made simply and inexpensively, using smartphones or digital cameras and voice recorders, and edited using free and simple software (Bow, 2017). Resources either collected or created for TELL in one context may be useful in others, for example, quizzes and tasks produced for TELL contexts can be used for vernacular literacy programs for children or adults, and videos used to demonstrate concepts such as kinship responsibilities or naming practices can serve various purposes for language and cultural programs more widely.

Identifying, collecting and creating appropriate resources is an important part of the process of course development, and such resources should be carefully archived (Nathan & Austin, 2014) to avoid the resources themselves becoming endangered (Bird & Simons, 2003). Storing materials only on websites or local servers is insufficient for ongoing access and sustainability.

The use of technology can be seen as removing certain protections by making things available online, however appropriate tools will have mechanisms for allowing and restricting access to certain users. Careful effort is required to manage the dual responsibilities of providing access and protecting intellectual property (Croft, Toussaint, Meakins, & McConvell, 2019). There are software solutions for storing linguistic and cultural knowledge with carefully thought-out processes and protocols for access, such as Mukurtu (Christen, Merrill, & Wynne, 2017) and Ara Irrititja (Hughes & Dallwitz, 2007). The resources contained in such tools could be shared for teaching and learning, however they are not designed for these practices.

TELL for Indigenous language use does not need to be at the ‘bleeding edge’ of technological advancement, but can draw on the experiences of bigger language groups in terms of both technology and pedagogy (Ward & van Genabith, 2003). The implication for designers is that awareness of the issues identified here will impact the ways in which development of TELL tools is approached.

4. Digital Language Shell

Consideration of all these issues informed the development of the Digital Language Shell project, designed to consider what was possible with limited resources and limited technical skills. Initially prompted by the lack of opportunities to study Indigenous languages at Australian universities, the goal was to provide a ready-to-use low-cost and low-tech template which draws on the affordances of TELL with considerable flexibility and customisability for different user groups in different contexts. The resulting Digital Language Shell provides a means of compiling and curating materials for language teaching and learning under appropriate Indigenous authority, without requiring much technical support or a large budget.

The project explored the possibilities of assembling a usable tool from off-the-shelf products. The selection process involved research into existing platforms aligned with a set of criteria, including cost (free or very low cost), customisability (of both appearance and functionality), flexibility (adding features for different purposes), simplicity (no coding required), support (free or low cost), multimedia options (supporting different file types for audio, video, text, image, etc), and Unicode-friendly (able to handle special characters).

WordPress was selected as the appropriate platform for the Digital Language Shell, providing a reasonable compromise of technology, budget and flexibility, while meeting all the established criteria. The huge range of plugins available for WordPress can be overwhelming, so the project involved experimenting with different options, reading reviews and checking compatibility. This process means that recommendations can be made, but users can explore further and choose alternate ones. Plugins for the Digital Language Shell include options for quizzes, multimedia playback, design options, survey and form types, security and backup features, discussion forums, payment processes, etc.

WordPress is not a learning management system, but there are several options for installing this functionality on the platform, including both paid and free plugins. For the purposes of this project, only free options were considered, in order to see what was possible for no expense. The idea was that a Digital Language Shell could give people the opportunity to get started in the TELL space, experiment with options, learn some skills, and deliver a customised product. For the Shell a free plugin called LearnPress was selected, which enabled the setting up of courses, units and lessons, plus a login
procedure (with an option to charge for courses), which was then populated with various content (text, media, links, etc.).

The resulting Digital Language Shell (see Figure 1) is a particular configuration of WordPress and a range of plugins, rather than a package that can be shared. A set of instructions has been written so that users can get started, with plenty of flexibility to explore alternative methods of delivery. All that is required to start is an installation of WordPress.org on a hosted server, then users can set up their site by selecting a theme for look and feel, which can be populated with appropriate local images, and plugins can be selected for additional functionality. Once a course is developed, administrators can easily make changes on the fly, rather than relying on external support, and can allow or restrict access to materials through various user profiles.

![Digital Language Shell](image)

**Figure 1.** Screenshots from Digital Language Shell and Bininj Kunwok online course

## 5. Users of Digital Language Shell

The first implementation of the Digital Language Shell was in collaboration with the Bininj Kunwok Regional Language Centre. Bininj Kunwok refers to a chain of related language varieties with around 2000 speakers across West Arnhem Land in the Northern Territory. Committee members from the language centre (who co-designed the curriculum with academic staff from Charles Darwin University) were not familiar with online language teaching tools, nor were they particularly interested in the technology itself. Their motivation was to share their language and culture with non-Indigenous professionals in the community. A pilot course of four units was developed and delivered to over 100 volunteer learners (Bow, 2017), and later this course was extended to a one-semester (12 week) course for university delivery online. The extension involved creating additional units and adding assessment tasks, to conform to the standard required of a university language course. This process led to producing more resources, strengthening teachers, creating student demand, and used Indigenous cultural concepts to build connections between speakers and learners (Bow, 2019). Evaluation of the tool is currently underway.

The second instantiation involved a very different linguistic ecology and academic context. The Muurrbay Aboriginal Language and Culture Co-operative in northern NSW works with a range of Aboriginal groups in language revitalisation. They offer in person classes Gumbaynggirr language, which has good documentation and recordings and a range of teaching materials. Seeking possibilities for an online option to serve the diaspora of Gumbaynggirr people scattered across the region and state, and lacking the capacity to develop their own course, they implemented the Digital Language Shell (Muurrbay Aboriginal Language and Culture Co-operative, 2019). Adapting the Bininj Kunwok course to suit local needs, they plan to use this as a model for teaching other languages of the region.

A number of other groups have expressed interest in using the Digital Language Shell for their purposes. In 2019 at the Puliima Indigenous Language and Technology conference, over 100 people
attended workshops to learn how about build their own courses at low cost and low technical demands. These examples highlight the need for a simple, inexpensive tool that can be used to support the aspirations of Indigenous people for language and cultural work for various purposes.

6. Conclusion

Technology can come to the assistance of Indigenous elders who are keen to share their languages and cultures. The variety of options, and heterogeneity of language contexts, may make it hard to know how to get started. Certain issues need to be addressed in the selection or design of TELL tools, including audience, costs, resources, archiving and access. The experience of the Digital Language Shell project is that it is possible to create a platform using free and open-source tools, with the flexibility to address various needs. The project does not focus on the technology itself but what it affords for Indigenous language groups, who can work collaboratively to design and develop course materials, always keeping in mind the aspirations and goals of the community members.

The Digital Language Shell is available as a model for others to use – it is not meant to be a solution to every problem, nor the best tool for the job, but it demonstrates what can be done little money and little technological skill, while involving Indigenous authorities in the process of course development. Respecting the authority of Indigenous language owners and increasing visibility of these languages can support the aspirations of Indigenous language communities and retain ownership of their knowledge and how it can be appropriately shared. Future directions can explore its application in other language contexts, how it can support the use and development of language documentation, incorporation of language varieties and cultural information, and how Indigenous groups use technology to support their aspirations. Such tools are practical and useful for Indigenous Australians beyond just language maintenance and revitalisation, but can also serve to enhance community well-being and education. Without the tools and means to implement such programs, these languages are unlikely to ever be made available.

Acknowledgements

The author would like to thank members of the Bininj Kunwok Regional Language Centre, and the project team at Charles Darwin University. The Digital Language Shell was initially funded by the Office of Learning and Teaching (SD15-5124), and the expansion of the course was supported by a Transdisciplinary & Innovation Grant (TIG842018) from the ARC Centre of Excellence for the Dynamics of Language. The author’s research is funded by an Australian Government Research Training Program Scholarship.

References


Learning Vocabularies through WeChat: A Case Study of Chinese Lower-proficiency Students

Chunlin LEI\textsuperscript{a*}, Yuyang CAI\textsuperscript{b}

\textsuperscript{a}Shanghai University of International Business and Economics, China
\textsuperscript{b}Shanghai University of International Business and Economics, China
lei.chunlin@suibe.edu.cn

Abstract: WeChat is an extremely popular social medium among Chinese students. This study designed and examined a WeChat-based environment in support of students’ English vocabulary learning. In a Comprehensive English class, a group of 21 Year-1 lower-proficiency college students were required to use new vocabularies to make sentences and post them to WeChat moments (friends circle) for others to share and learn. It is hoped students’ WeChat learning activities can consolidate their vocabulary learning. The data sources of this study included students’ WeChat logs, vocabulary tests during mid and end of instruction, and interviews with focus groups of students. The results showed all students participated in the online vocabulary learning tasks and their achievement on vocabulary tests maintained over time. However, students’ levels of engagement varied. The interview data revealed a number of socio-cultural factors contributing to (or undermining) the students’ engagement and improvement, for example, good personal relationship, a growth mindset and a need to show off (or oppositely, a lack of trustworthiness and a concern for losing face). The implications of the study are discussed at the end of the paper.

Keywords: WeChat, vocabulary learning, lower-proficiency students, engagement

1. Introduction

The importance of vocabulary learning and teaching in English-as-a-foreign-language (EFL) context has long been recognized (Nassaji & Tian, 2010; Rassaei, 2020), and the case is even more pressing for lower language proficiency students. These students need to start with vocabularies and to be more exposed to foreign language environments in order to build up confidence and step toward success (Zimmerman & Schunk, 2011). With the widespread of information technology in education, many technology-based tools and methods have been utilized in vocabulary learning and achieved fruitful results (see meta-analysis, Abraham, 2008; Bax, 2011).

WeChat, known as “微信” in Chinese, is an Internet-supported social medium, which has garnered over 1150 million monthly active users by June 2019 (Tencent, 2020) and is extremely popular among Chinese people. WeChat has deeply changed people’s lives (Gan & Li, 2018); it has been used to facilitate communication with friends, understand various kinds of news (Gan & Wang, 2015), and support teaching and learning (Wang et al., 2017; Wang, 2018). Students’ playing with WeChat through smartphones seems ubiquitous on university campuses, and teachers often find it a distraction to classroom teaching and learning; however, it is neither practical nor feasible to forbid students from using phones and a more constructive idea may be to exploit opportunities of WeChat in support of students’ learning, be they in formal (in class) or informal (out of class) learning settings. It is against such background, this study intends to design and examine a WeChat supported learning environment through which students can make sentences, interact with one another, and improve their vocabulary and language learning.

2. Literature Review

Mobile-assisted language learning has emerged to be a new research strand in recent decades (Warschauer & Kern, 2000). In its early stage, the most frequently used function of mobile phones is...
message texting, which has been used to assist students in their vocabulary learning. With technology development and popularity of smartphone apps, new approaches of learning with the help of internet and apps have been widely adopted (Lu, 2008). In learning vocabularies, researchers have rigorously designed and investigated mobile-based vocabulary learning activities (Abraham, 2008; Stockwell, 2007, 2010). For example, Stockwell explored the impact of smartphone on Japanese students’ vocabulary learning, and focused on a comparison of time spent on the task via mobile phones or computers (2007) as well as learners’ usage patterns of mobile platforms (2010) respectively. The results indicate that mobile phones have the potential to enlarge students’ vocabularies through some sophisticated vocabulary learning activities. In mainland China, Yang (2012) explored mobile technology in college English vocabulary learning and proposed that the method had transcended the limits of time and space and performed better than the traditional method. In a more relevant study, Wang (2018) has taken advantage of a quasi-experimental design to investigate the use of WeChat group and English Learning Apps in support of students’ vocabulary learning. The results showed different achievement groups had different preferences towards mobile learning and their performances on vocabulary assessment were different. Although mobile assisted language learning has generally demonstrated educational benefits in research studies, the mechanisms behind the benefits are still under-investigated and studies shall not only focus on the affordances of technology but also the constraints (Reinders & Stockwell, 2017).

WeChat is a widely adopted social medium across China and even all over the world; however, most of the studies on WeChat have focused on users’ behaviour changes (Peng, Zhao & Zhu, 2016) and premised on gratification theory (Gan & Li, 2018), and there is a dearth of research on the use of WeChat in support of academic learning. In view that WeChat is not specifically designed for classroom learning; application of WeChat in vocabulary learning may bring about both opportunities and challenges. In addition, students of lower proficiency may lack skills and motivation, and employ strategies such as avoidance and self-defence in learning (Zimmerman & Schunk, 2011). To what extent they may feel motivated to use WeChat to learning needs further exploration.

This paper addresses two research questions: (1) how do the students participate in the WeChat-supported vocabularies learning activities and perform in the subsequent vocabulary tests; (2) what are the factors that affect students’ engagement in WeChat vocabulary learning environment?

3. Research Method

3.1 Research Context

This study was conducted in a Sino-British joint educational program at a business-and-economics-oriented university in Shanghai, China. The learners of the Program were lower-proficiency students because they were not formally recruited through China’s Matriculation System (or Gaokao). Most of them failed the Chinese Gaokao, and then relied on the “3+1” joint Program (3-year study in Shanghai and 1-final-year study in Britain) to continue their study in higher education. Upon successful completion of the Program, the student could obtain a BA degree from the British university. Participants of this study included a group of 21 Year-1 students (M=10, F=11), aged from 17 to 20. The teacher is a young, novice teacher with 2-year teaching experience.

3.2 Instructional Design

The students were registered in a course entitled Comprehensive English, a foundational course aiming at strengthening students’ ability in vocabulary, grammar and English reading literacy. The course was delivered in a 90-minute lesson everyday from Monday to Friday across an 18-week length semester. As vocabulary learning was one of the primary tasks of the course and the students used their smartphones quite often in and outside the class, the teacher required the students to make use of WeChat as a supplementary tool for vocabulary learning. Specifically, students were equally (except for one group) divided into five groups and each member of the group was supposed to use three new vocabularies that he/she learnt from the class or textbook to make sentence(s) and post those sentences in their WeChat moments (friends space). Since WeChat moment is an open space where all friends
(including classmates, teachers, even parents and relatives) have access to, students were therefore, encouraged not only to complete the sentence-making tasks but also to read and give comments to their classmates’ sentence-making. Through these online activities (usually after class), it was hoped that students could better understand the meaning of the words, the contexts of the words when used; and more importantly, they could learn from one another.

3.3 Data Collection

The data sources of this paper included (1) students’ posts in the WeChat moments (sentence-making, comments, responses, click of liking – a way to show appreciation to others’ posts, etc.) throughout the semester, which provided a holistic picture of students’ engagement in the environment; (2) two vocabulary tests during mid (Week 9) and end (Week 18) of the semester; each test consisted of 60 vocabulary questions which were carefully selected and designed by the teacher and the researcher in line with the textbook, students’ WeChat posts, and degree of difficulty (e.g. Word List, Coxhead, 2000); and (3) focus group interviews with the students of high (n=4), middle (n=4), and low (n=4) engagement in WeChat environment (selection based on students’ vocabulary tests scores, their WeChat logs, and the instructor’s observation); the interview transcripts helped to shed light on the socio-cultural factors that affected students’ engagement in the WeChat environment.

4. Findings

4.1 Students’ Participation in the WeChat-based Learning Activities

The mobile phone logs were kept to track the students’ participating behaviors in the WeChat environment. As WeChat is used more as a social medium and the posts are much diversified (including pictures, texts, hyperlinks, etc.), this paper only focuses on the entries relating to vocabularies and sentence-making. Since sentence-making was a required assignment for students registered in the course, all the students did it accordingly. However, their posts vary in number, length and quality. Some students deliberately put three vocabularies into one sentence thus made only one entry each time; while other students preferred to use one or two vocabularies in one sentence and thus made two/three sentences/entries each time. In total, 21 students made 571 sentences, with a Mean value (M) of 27.19 and a Standard Deviation (SD) of 8.84.

One big incentive to use the tool was to promote students’ online interaction or peers’ comments and responses pertaining to sentence-making. The dataset showed students received 99 pieces of peers’ comments (such as “A good sentence”, “awesome, buddy”, “The clause is wrong”), excluding those of the instructor’s. The M and SD were 4.71 and 5.58 respectively. Similarly, students gave out 99 comments, and the M and SD were 4.71 and 5.26, respectively. These results suggested students generally had limited interactional behaviours in the WeChat vocabulary learning environment and they differed greatly in participating in the learning activities as well.

4.2 Students’ Performance in Vocabulary Tests

Two vocabulary tests were administered during Week 9 and Week 18 respectively. The results are shown in Table 1. A paired test showed there was no significant difference between the two tests, $T (20) = 1.49, P=0.15>0.05$. However, students’ performance varied dramatically across individuals if we put the M, SD, highest and lowest scores into consideration.

Table 1: Students’ Performance in Two Vocabulary Tests

<table>
<thead>
<tr>
<th></th>
<th>Number of participants</th>
<th>Mean (Max=60)</th>
<th>Standard deviation</th>
<th>Highest score</th>
<th>Lowest score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week-9 test</td>
<td>21</td>
<td>29.1</td>
<td>14</td>
<td>51</td>
<td>11</td>
</tr>
<tr>
<td>Week-18 test</td>
<td>21</td>
<td>26.5</td>
<td>10.6</td>
<td>49</td>
<td>10</td>
</tr>
</tbody>
</table>
4.3 Factors Affecting Students’ Engagement and Vocabulary Achievement

As explained in data collection part, 12 students representing high, mid and low achievers in the group were invited semi-structured interviews, which delved into their feelings, experiences and behaviours when working on the WeChat environment. The transcripts were analysed using a thematic approach; and the themes or factors contributing to students’ sustained engagement in vocabulary learning activities (as well as the factors hindering students’ participation and efforts) were identified as follows.

4.3.1 Factors Promoting Student Engagement

**Good personal relationship.** “Relationship” or Guanxi is an element deeply rooted in Chinese culture, which is usually interpreted as a dyadic connection based on mutual interest and benefit. Usually, people are willing to offer a favour or help to those with whom they share good relationships. Ten of the 12 interviewees, explicitly or implicitly, expressed their willingness to participate in the online assignment/activities because they enjoy some good and sustained relationship with the class instructor (Miss Anny). For example, “Anny is gentle and nice to me [Excerpt #1, ZM]”; “She is patient, just like my big sister, I sometimes leave other teachers’ assignments undone, but I never failed hers [Excerpt #2, XJT]”. In addition to student-teacher relationship, good student to student relationship also encouraged engagement. “I only follow (give feedback to) friends of good relationship, although all classmates are within WeChat moment, we have different relationship ... [Excerpt #3, LY].”

**A growth mindset.** Interview results also revealed students’ beliefs about learning, especially their strategies adopted to cope with difficulties and challenges are important predictors of their engagement and improvement. A positive mindset helps students seek opportunities to learn and make learning an interesting endeavour. For example, Student WQH, a relative high-achiever among the group of lower-proficiency students, mentioned “In choosing vocabularies to make sentences, I will deeply reflect on what I have done the whole day or recently; and fit the words in sentences describing my life or the current scenarios [Excerpt #4]”. Another student ZM, who enthusiastically interacted with her classmates on the WeChat, explained that “… I don’t mind Prof. ZH (teacher of another course and happens to be ZM’s WeChat friend) pinpointed my mistakes in sentence-making in the WeChat moment... no shame at all in the (online) public place. Instead, I feel grateful to him...People who know me will see I am making progress bit by bit ... [Excerpt #5]”. Obviously, a growth mindset, or accepting mistakes and seeing it as a springboard for stretching her existing abilities, motivated ZM for her active participation.

**A pragmatic showing-off need.** A typical feature of WeChat is its openness and the interview data indicated that showing off one’s abilities or efforts in public could be a strong motivator for students’ participation and interaction in WeChat moments. The case is especially true with students possessing relatively good academic records, for example, WQH mentioned “I often read my classmates’ sentences, and it is not difficult for me to spot out their mistakes, mainly grammar…I feel aha (being proud)...I can tell them now [Excerpt #6]”. Such a sense of pride (stemming from identifying others’ mistakes) is coupled with a need to make a show of his own ability as WQH continued in his remarks “… I hope what I wrote could act as the model of the class [Excerpt #7]”. In another case, Student WSJ who is only a mid-achiever in the class, also expressed the need to show-off, not amid her teachers or classmates but her parents and relatives, enhanced her WeChat activities. “My parents used to complain not knowing what I was doing on campus … when I publish sentences on WeChat, they quickly realize I am studying (a decent endeavour)...that is important! [Excerpt #8]”.

4.3.2 Factors Undermining Student Engagement

**Lack of trustworthiness.** Although WeChat acts as a social medium and the moments are supposed to contain circle of “friends” only, the actual situation is more often that one “added or enlisted” many people who are not necessarily his/her true friends. Even within the classroom boundary, some students still feel unsafe to give feedback to others in public. “Although we are classmates and WeChat friends, isn’t it embarrassing to give comments to someone with whom I am not very familiar? [Excerpt #9, WZR]”. Other manifestations of untrustworthiness include students’ scepticism about using WeChat as
a learning tool or their doubt on the classmates’ abilities. “WeChat is a place where you show food, travel, and your life … and relax…how can you share some strange stuff (English sentence-making) in the WeChat moment? [Excerpt #10, LY]”; “As you know, we joint program students are weak students; I don't think it fruitful to read the sentences made by the students…[Excerpt #11, LHY].”

A concern for losing face. Face value is an important concept in the Eastern culture and Chinese people usually have a strong consciousness towards face. Being afraid of face-losing in public is a big factor that hindered students’ WeChat activities. For example, Student ZT said she only finished her task and never commented on other students’ posts because she could not afford to lose face, “I am not sure what I say is correct or not, if I say something wrong, that’s a loss of face in public [Excerpt #12, LHY]. In addition, Student SRY even avoided a click of “liking” (an easy usual behaviour showing agreement or appreciation on WeChat) because he was afraid if he clicked liking while there were problems in the sentence-making, “it would still look a bit foolish” and the best thing for him to do is to “do nothing” [Excerpt #13].

5. Discussion

This study took advantage of a popular social medium, WeChat, to support vocabulary learning among a group of lower-proficiency students in China. The results showed that the students were able to participate in the platform tool and accomplish their sentence-making tasks. In the design, there demonstrated some novelty effects of technology as some interviewee students expressed they “had never thought of using WeChat to learn English vocabularies and felt intriguing". Such phenomena are quite common when technology is first introduced to students in support of learning. This study made some good efforts to facilitate students’ vocabulary learning. Sentence-making is perhaps a good strategy to use. As we can see in the study some students “associated the sentence-making with their life scenarios and contexts and showed them to the public”, which coincides with principles of foreign language learning; for example, “learning in authentic situations” and “meaning-based, output-oriented learning (Wen, 2015)”. However, it is still difficult to draw a conclusion regarding the effects of the WeChat vocabulary learning environment. The affordances and constraints co-exist. For example, students did two vocabulary tests and the results were quite similar, in a way it could be interpreted as students’ performance sustained over time (mid- and end of instruction); however, students’ correction rates of vocabulary tests were still low (46.9% and 44.2% respectively) and the score even dropped a bit at the second time. Therefore, future longitudinal studies or quasi-experimental design involving a comparison group may provide more convincing pictures of the learning effects.

This study using WeChat did overcome the traditional limits of time and space in learning (Yang, 2012); however, each tool has certain functions that may work or fail depending on their alignment with the pedagogical or socio-cultural dynamics in context. With WeChat’s pervasiveness, convenience and a public space (moments), students have more opportunities to be exposed to learning materials and learning could informally happen even during social interactions. However, such practice worked with some students but not all the learners. There are a lot of contextual factors coming into play, for example, the relationship between teacher to students and students to students, students’ beliefs about learning and about themselves, and their needs to show-off or concern about losing face. All these factors, if not handled properly, are very likely to constitute the constraints of mobile assisted technology (Reinders & Stockwell, 2017). For example, students in this study showed uneven participation and engagement patterns. Some students avoided WeChat-based learning activities because of insufficient confidence and trust. To make technology-enhanced environment more effective for learning, it is suggested fostering a democratic, trustworthy atmosphere in class or a learning culture is essential (Lei & Chan, 2018).

This study has shed some lights on how to deal with students of lower proficiency. These students are usually featured by handicapping learning strategies and low motivation (Zimmerman & Schunk, 2011). It is therefore reasonable to tap every opportunity to motivate them to learn. Utilizing an existing technology (smartphone app) to help students learn is better than suppressing students from using the technology (sometimes a distraction to classroom learning). However, when the openness of social media is intertwined with learning, we need to be very careful in the design. For example, the weak students can be very vulnerable and sensitive to losing face in public (WeChat moments or friends
Therefore, the class instructor needs to provide scaffolds to these students, for example, modelling of making mistakes in public, in order that they develop a growth mind-set and understand making mistakes is commonplace in learning and a necessary path to success and improvement. On the other hand, the relatively good students in the group, when equipped with some showing-off needs, should be properly guided so that their efforts can be encouraged and sustained, and they can help others to grow.

The limitations of the study include small sample size, a lack of explanations of the instructor’s role in the study, etc. Future research will address these issues and provide more holistic landscapes of the socio-cultural dynamics.

References


Multi-Channel CNN-BiLSTM for Chinese Grammatical Error Detection

Lung-Hao LEE, Yuh-Shyang WANG, Po-Chen LIN, Chih-Te HUNG & Yuen-Hsien TSENG

Department of Electrical Engineering, National Central University, Taiwan
Graduate Institute of Library and Information Studies, National Taiwan Normal University, Taiwan
*samtseng@ntnu.edu.tw

Abstract: In this paper, we proposed a Multi-Channel Convolutional Neural Network with Bidirectional Long Short-Term Memory (MC-CNN-BiLSTM) model for Chinese grammatical error detection. The TOCFL learner corpus is adopted to measure the system capability of indicating whether a sentence contains errors or not. Our model performs better than a previous CNN-LSTM model that reflects the effectiveness of multi-channel embedding representation.

Keywords: Chinese as a foreign language, grammatical error diagnosis, deep neural networks

1. Introduction

Chinese as foreign language (CFL) learners usually make various kinds of grammatical errors, such as missing words, redundant words, incorrect word selection, or word ordering error, during their language acquisition and production process. An automated system able to detect such errors would facilitate the learning and teaching of CFL. Previous Chinese grammatical error detection approaches were based on linguistic rules (Lee et al., 2013), machine learning classifiers, or their hybrid methods (Lee et al., 2014). Deep neural network (i.e., deep learning) approaches have been widely applied recently and achieve dominating results in many natural language processing tasks. This trend motivates us to explore deep neural networks to detect errors written by CFL learners.

This study describes our proposed Multi-Channel Convolutional Neural Network with Bidirectional Long Short-Term Memory (MC-CNN-BiLSTM) model for Chinese grammatical error detection. The TOCFL learner corpus (Lee et al., 2018) is used to evaluate the performance. Compared with previous approaches on the same dataset, our proposed method achieved better F1-score which takes into account both detection precision rate and recall rate at the same time.

2. Multichannel CNN with Bidirectional LSTM (MC-CNN-BiLSTM)

Figure 1 illustrates our Multi-Channel Convolutional Neural Network with Bidirectional Long Short-Term Memory (MC-CNN-BiLSTM) model for Chinese grammatical error detection. The model has two main parts: 1) the multi-channel embedding representation and 2) a CNN along with a Bidirectional LSTM network (CNN-BiLSTM). In this model, an input sentence is represented as a sequence of words. The embedding vector of each word is pre-trained from a large corpus using different embedding techniques, such as Word2vec (Mikolov et al., 2013), GloVec (Pennington et al., 2014), and ELMo (Peters et al. 2018). Through different delicately designed training, the embedding vectors learn to become better semantic representations of their words in different ways. Therefore, instead of adopting only one single embedding technique for semantic representation, the multi-channel representations adopting multiple embedding sources as multi-channels are used in our model.

In each channel, a convolutional layer with $k$ different convolution filters is used to extract $k$ feature maps, each contains a kind of local $n$-gram (a consecutive of $n$ word embeddings) features. The max-pooling layer is then used to aggregate the local $n$-gram features from each feature map to keep the most salient information corresponding to each input word. The obtained feature vectors are fed to the sequential BiLSTM layer to capture long-distance (global) features among the input words. In the BiLSTM, the backward LSTM is a reversed copy of the forward LSTM, so that we can take full advantage of the forward and backward word sequence information from the input sentence. The output of the BiLSTM denoting the input sentence representation from each channel is concatenated to form a
long vector, which is fed to a neural network layer with the final softmax activation function for binary classification signifying whether the input sentence contains grammatical errors or not.

During the training phase, if a sentence contains at least one grammatical error judged by a human, its class is labeled as 1, and 0 otherwise. All the sentences with their labeled classes are used to train our MC-CNN-BiLSTM model to automatically learn all the corresponding parameters in this model. To classify a sentence during the testing phase, the sentence unseen in the training phase goes through the MC-CNN-BiLSTM architecture to yield a value corresponding to the error probability. If the probability of a sentence with class 1 (i.e., with errors) exceeds a predefined threshold, it is considered as true as an erroneous sentence, and false otherwise.

3. Experiments and Evaluation Results
The experimental data came from the TOCFL learner corpus (Lee et al., 2018), including grammatical error annotation of 2,837 essays written by Chinese language learners originating from 46 different mother-tongue languages. Each sentence in each essay is manually labeled. The result is that a total of 25,057 sentences contain at least one grammatical error, while the remaining 63,446 sentences are grammatically correct (an unbalanced distribution with 28.31% sentences having grammatical errors). Five-fold cross validation evaluation was used to measure the performance.

For Word2vec, GloVe, and ELMo embedding representations, the whole Chinese Wikipedia (zh_tw version on Dec. 24th, 2019) was firstly segmented into words and then the segmented sentences were used to train 300 dimensional vectors for 849,217 distinct words. For convolution operation, the number of filters was \( k \)=300 and their length was 3. The loss function was categorical cross-entropy and the optimizer was Adam. The number of training iteration (i.e., epochs) was set to 3 to learn the CNN-BiLSTM network parameters. If the error probability of an input sentence exceeds 0.2, it was considered as an erroneous sentence.

The following methods were compared to show their performance. (1) CNN-LSTM (Lee et al., 2017): this method integrated CNN with LSTM. According to their suggestions, Word2Vec embedding was used and the hyperparameters were set up as follows: The number of filters was 300 and their length was 3, epochs were 5, and the threshold is 0.3. (2) MC-CNN-BiLSTM: this is our proposed
model for Chinese grammatical error detection. We also do some ablation experiments, i.e., comparing the performance using different number of channels in terms of individual embedding combinations.

Table 1 shows the results. The single-channel MC-CNN-BiLSTM was degraded as CNN-BiLSTM, which was still slightly better than the CNN-LSTM (Lee et al., 2017). It reveals that bidirectional LSTM in the sequence layer can enhance the performance. Comparing our MC-CNN-BiLSTM model with the CNN-LSTM, all our models has the improvement in terms of F1-score no matter which number of channels was used. In our observation, an increasing number of channels will achieve better performance if different embedding representations were used in a proper combination. It is noted that due to the large number of testing examples (17700=(25057+63446)/5), the small number of improvement (over 0.01 absolute improvement or over 3.72% relative improvement) is still statistically significant.

Table 1: Evaluation on Chinese grammatical error detection.

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN-LSTM (Lee et al., 2017)</td>
<td>0.3812</td>
<td>0.6544</td>
<td>0.4808</td>
</tr>
<tr>
<td>MC-CNN-BiLSTM (our model)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single-channel</td>
<td>Word2vec</td>
<td>0.3386</td>
<td>0.8656</td>
</tr>
<tr>
<td>Dual-channel</td>
<td>Word2vec + GloVe</td>
<td>0.3667</td>
<td>0.7801</td>
</tr>
<tr>
<td>Triple-channel</td>
<td>Word2vec + GloVe + ELMo</td>
<td>0.3669</td>
<td>0.7845</td>
</tr>
</tbody>
</table>

4. Conclusions

Chinese grammatical error detection is a very hard problem such that even MS Word has no such function for Chinese (but has this function for English for many years). This study describes the MC-CNN-BiLSTM model for Chinese grammatical error detection. We use the TOCFL learner corpus to demonstrate the system performance. Our system achieved an improvement of F1-score 0.4987 for predicting whether a given Chinese sentence contains any grammatical errors or not, which roughly corresponds to a half of input sentences were judged correctly under the unbalanced error distribution.

Acknowledgements

This study was partially supported by the Ministry of Science and Technology, under the grant 109-2410-H-003-123-MY3, 108-2218-E-008-017-MY3, and 106-2221-E-003-030-MY2.

References


Development of Japanese Dictogloss Learning Support Environment for Pronunciation Learning of Japanese Speech

Satoru KOGUREa*, Hiroki HAKAMATAa, Yasuhiro NOGUCHIa, Tatsuhiro KONISHIa, Makoto KONDOa & Yukihiro ITOHb

aFaculty of Informatics, Shizuoka University, Japan
bShizuoka University, Japan
*kogure@inf.shizuoka.ac.jp

Abstract: In this paper, we describe the extension of the existing pronunciation evaluation framework of Japanese dictogloss support environment. To evaluate the pronunciation in the existing environment, it was necessary for the teacher to manually create a dictionary for recognition that include incorrect pronunciation. In addition, there were few variations in the dialogue part that indicated an error in the reproduction sentence of the collaborative learner agent and the sentence of pronunciation evaluation feedback. To solve these problems, we improved the system. We evaluated the system we created and obtained the results that suggest that the proposed method is effective.

Keywords: Dictogloss, pronunciation evaluation, language learning support, Japanese learning

1. Introduction

Dictogloss is an activity in language learning advocated by Wainryb (1988, 1990). An advantage of dictogloss is that it can teach speaking skills in addition to reading, writing, and listing. There are many practical reports on its application in the language-learning field (Snoder & Reynolds, 2019, Ajmal et al., 2020). However, in dictogloss, the learner needs a teacher and a collaborative learner. Therefore, it is difficult to use this activity when alone. To address this problem, we have developed a dictogloss system (Kondo et al., 2012; Tashiro et al., 2013, Kogure et al., 2015, 2016, 2017, 2018) that does not require teachers and collaborative learners. Our dictogloss environment has two agents: a collaborative learner agent (CLA) and a teacher agent (TcA). However, our dictogloss environment faces the following has three problem: (P1) the cost of generating a speech recognition dictionary for pronunciation evaluation is high. (for teaching material creators), (P2) there are limited interaction patterns that learners can experience. (for learners), and (P3) pronunciation evaluation feedback is binary (for learners). The purpose of our research is to solve these three problems, and in this report, we propose a method to solve them. We also implement a system that utilizes the method. In addition, we evaluate the system we created.

2. Dictogloss Learning Environment

Dictogloss is a learning activity advocated by Wainryb (1988, 1990). It contains four activities:

(Act.1) The teacher reads a short text aloud (including some sentences).
(Act.2) Each learner listens to the text while taking notes.
(Act.3) Each learner reconstructs the original text while consulting with another learner.
(Act.4) The teacher reviews the reconstructed sentences in text and provides feedbacks to learners

There are many class practice reports that utilize dictogloss in actual language classes. Consider providing information and communication technology support for dictogloss learning activities. First,
in Act. 1, we can realize a system in which the learner can freely reconstruct by recording the teacher’s speech. Next, in Act. 2, we can prepare an environment in which the learner writes down the words heard. However, in Act. 3, it is necessary to build an environment in which it is possible to interact with collaborative learners. Therefore, this step is difficult to realize. Also, Act. 4 requires the same high level of difficulty in realization as in Act. 3 because it requires analysis of the learner’s utterance in Act. 3. We have devised a method that simplifies the interaction with collaborative learners in Act. 3 and the method of analyzing the learner's input in Act. 4. We constructed a Japanese dictogloss learning support environment that supports from Act. 1 to Act. 4 according to the method (Kogure et al., 2018).

In Act. 2, the learner utters the sentence reconstructed by himself or herself. With this function, if the learner does not enter the reconstructed sentence correctly (or if it contains an unexpected error), the reconstructed sentence cannot be recognized correctly. Therefore, the learner can utter the reconstructed sentence only when she or he inputs the reconstructed sentence within the range of the assumed error. In Act. 3, the learner compares the reconstructed sentences input by himself or herself and the collaborative learner. The learner specifies different morphemes and points out errors to the collaborative learner. At that time, the learner can ask "Was there WORD in Sn?". The system recognizes the pointed utterance by speech and then identifies the word for which the error was indicated. The system generates a click event of the morpheme button corresponding to that word. In Act. 4, the system analyzes the learner's voice recognition results collected in Act. 2 and 3 and evaluates pronunciation (Kogure et al., 2018).

3. Improvement of Pronunciation Evaluation

The teaching material creator (1) generates a speech recognition grammar and a dictionary with correct pronunciation, and (2) a pronunciation dictionary that includes erroneous pronunciations that are easily mistaken by a native speaker. They have to generate manually, and it especially takes time to complete (2) task for them. For example, it took an average of about 500 seconds for the author to create the incorrect pronunciation dictionary (for Korean native language) for 3 lessons out of 22 lessons prepared. We expect it to be more expensive to be created by someone who is not familiar. In addition, we judge that there is a possibility that the points to be corrected may be omitted or the quality of the created dictionary may be deteriorated by manual correction. In addition, since phonetic transcriptions that are expected to be incorrect differ depending on the mother tongue, it is necessary to manually correct each time the mother tongue changes. Therefore, when a mother tongue is specified, a framework that can automatically generate an erroneous pronunciation dictionary that considers pronunciation prone to error in the mother tongue is incorporated. The system only allows the learner to utter “Was there WORD in Sn?” and cannot use other utterances. Therefore, we increase the number of variations of dialogue that learners can experience. The actual pronunciation evaluation is performed on a word-by-word basis, but the confidence measure of pointing out feedback uses sentence confidence measure. The system does not look at word confidence measure. Therefore, we improved the usage of speech recognition. The system estimates the confidence measure for each word using the speech recognition results up to 5 best results.

4. Experimental Evaluation

For lessons 1–3, one of the authors manually created a speech recognition dictionary containing incorrect pronunciations. The generation time was 303 seconds for Lesson 1, 488 seconds for Lesson 2, and 741 seconds for Lesson 3. On the other hand, it took less than 1 second for all three lessons to automatically generate the dictionary. This result shows that P1 was solved.

Regarding the part that points out the error in CLA in Act. 3, we let four Japanese people use utterances that point out the error due to voice utterance in the preceding and the proposed system. After that, we acquired the goodness of dialogue by increasing the number of styles that can be spoken. We prepared two target lessons. Considering the order effect, we changed the order of the lessons used and the order of the system used by all four subjects. We asked the four subjects to evaluate whether each system was easy or difficult to utter to point out a mistake, using 1 to 5 points (1 meaning that previous system is very good, and 5 that the proposed system is very good). We obtained 2.25 points in the
preceding system and 3.75 points in the proposed system. This result suggests that increasing the variation of utterances that point out errors in Act. 3 may improve the ease of pointing out by learners.

We made the same four subjects input the reconstructed sentence prepared here in Act. 2. Then, we read the reconstructed sentence to the subject with the pronunciation prepared here (both correct and incorrect). All four of them read with the pronunciation presented here (all the same). After that, we made them read the text of the indication of the correctness judgment of the feedback of Act. 4. In this experiment, the preceding system and the proposed system were used in order. Considering the order effect, two subjects used the preceding system first, and the other two used the proposed system first. After use, we asked the learners to evaluate whether the utterance pointed out was useful for language learning (role-playing a learner who is likely to make a specified pronunciation error) on a scale 1-5 (1 meaning that previous system is very good, and 5 that proposed system is very good). We obtained 3.50 points in the preceding system and 4.75 points in the proposed system. This result suggests that increasing the variation of utterances that point out errors in Act. 3 may improve the ease of pointing out by learners, implying this result suggests that the proposed system could give a more appropriate feedback depending on the pronunciation of each word than the previous system.

5. Conclusion

In this research, we construct three improvements in the pronunciation evaluation framework of the existing Japanese dictogloss support environment. The first is the automatic generation of an erroneous pronunciation dictionary, the second is an increase in variability in utterances that point out mistakes to CLA, and the third is an increase in feedback variations for erroneous pronunciations. We evaluate each improvement and confirmed the usefulness of the system. It is necessary to conduct an evaluation experiment for international students who are learning Japanese. Also, we are currently planning to develop a dictogloss support environment for English.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 17K00483.

References


A Comparative Study on the Translation Quality of Specialized and General Machine Translation Outputs

Xiaohan YU¹, Guangming YU¹, Zhiyang, LIU² & Chunping ZHENG¹*

¹Beijing University of Posts and Telecommunications, China
²China National Institute of Standardization, China
*zhengchunping@bupt.edu.cn

Abstract: This study compares the translation quality among two general machine translation systems, Google Translate and Youdao Translation, and one specialized machine translation system, Smart Translation System for Standards (STSS). This pilot study explored the effectiveness of STSS and its translation errors, providing suggestions for language learners, educators, and users of machine translation.

Keywords: Machine Translation, Machine Translation Output, Evaluation Methods, BLEU, Translation Quality

1. Introduction

Apart from general translation platforms like Google Translate, research institutes and enterprises have developed domain-specific MTs to translate texts in a specified technical domain. Unlike general translation systems, however, there are few studies investigated the effectiveness of specified MTs when translating domain-specific documents. The comparative study explores the translation quality and errors of translation output of three machine translation platforms, namely, STSS, Google Translate and Youdao Translation. Based on our research, we provided suggestions for translation quality evaluation and language learning.

It is guided by two research questions:
1. What are the differences in performance of three platforms when translating documents about standards?
2. What are the differences of errors in translation outputs between general and specified machine translation platforms?

2. Research Methodology

2.1.1 The Translation Platforms in the Study and Source Texts

STSS was developed by China National Institute of Standardization (CNIS) as a specified MT for conducting bilingual translation (Chinese and English) of documents about national and international standards. Google Translate is one of the most popular MTs globally while Youdao Translation is a China-based and one of the most used MTs in China. The little difference is that STSS has its own translation memory (TM) which makes it a domain-specific translation engine.

We selected six documents which were included in the STSS platform TM (in-TM), and six not included in the TM (out of-TM). In-TM source texts consist of 6408 Chinese characters and 4532 English words while out of-TM texts consist of 6203 Chinese characters and 5107 English words. That’s all the source text used in the study.

2.1.2 Automatic Evaluation (AE) and Human Evaluation (HE)
We applied BLEU (Bilingual Evaluation Understudy) as our automatic evaluation tool. The human references of translation, which is required by BLEU metric, are all well-translated and published documents provided by CNIS. Quality Evaluation Code for Localization Translation and Desktop Publishing is a human evaluation standard issued by the Translators Association of China. Based on the standard, we classified the translation errors. After AE and HE, we are able to compare the performance of the three MT platforms. The comparison was carried out by two professional translators separately and cross-checked to ensure the accuracy.

3. Results and Conclusion

3.1.1 AE-based Comparison

Table 1 and 2 show the BLEU scores in two scenarios. For in-TM texts, STSS platform presents a higher result than Youdao Translation and Google Translator for two-way translation. For out of-TM texts, the three platforms do not show significant difference in scores.

<table>
<thead>
<tr>
<th>Source Text</th>
<th>Language</th>
<th>Yuodao Translation</th>
<th>Google Translate</th>
<th>STSS Platform</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-TM</td>
<td>Chinese-English</td>
<td>0.27</td>
<td>0.24</td>
<td>0.31</td>
</tr>
<tr>
<td>In-TM</td>
<td>English-Chinese</td>
<td>0.17</td>
<td>0.16</td>
<td>0.29</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source Text</th>
<th>Language</th>
<th>Yuodao Translation</th>
<th>Google Translate</th>
<th>STSS Platform</th>
</tr>
</thead>
<tbody>
<tr>
<td>Out of-TM</td>
<td>Chinese-English</td>
<td>0.25</td>
<td>0.23</td>
<td>0.27</td>
</tr>
<tr>
<td>Out of-TM</td>
<td>English-Chinese</td>
<td>0.16</td>
<td>0.18</td>
<td>0.19</td>
</tr>
</tbody>
</table>

3.1.2 HE-based Comparison

Table 3 and 4 show the error distribution of the two scenarios. Term error is the most frequent type for all three MTs followed by expression error. STSS did much better for in-TM texts while nearly the same as the other.

<table>
<thead>
<tr>
<th>Error Distribution</th>
<th>Google Translate</th>
<th>Youdao Translation</th>
<th>STSS Platform</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slight Errors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expression</td>
<td>12</td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td>Language Style</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Grammar</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Specific Symbol</td>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Minor Errors</td>
<td>Term</td>
<td>28</td>
<td>26</td>
</tr>
<tr>
<td>Major Errors</td>
<td>Mistranslation</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Error Distribution</th>
<th>Google Translate</th>
<th>Youdao Translation</th>
<th>STSS Platform</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slight Errors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expression</td>
<td>14</td>
<td>18</td>
<td>14</td>
</tr>
<tr>
<td>Language Style</td>
<td>5</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Grammar</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Specific Symbol</td>
<td>2</td>
<td>2</td>
<td>6</td>
</tr>
</tbody>
</table>
STSS performed generally acceptable and significantly better translation output than general machine translation platforms for the in-TM texts. However, its translation quality was nearly the same as that of general MTs for the out of-TM texts. This may due to the translation memory which makes the STSS sharper for translation of standards. We can safely say that the scale and quality of parallel corpus of translation memory is critical to improve the effectiveness of a domain-specific MT. It would be helpful for language learners to choose a domain-specific MT when learning terminologies of the technical field. From the perspective of error distribution, post-editors should pay special attention to term translation when using MT platforms. As we only used two methods for quality evaluation, the assessment needs to be refined by more comprehensive metrics.

Acknowledgements

This research is supported by the Fundamental Research Funds for the Central Universities (2019XD-A04) and the Project of Discipline Innovation and Advancement (PODIA)-Foreign Language Education Studies at Beijing Foreign Studies University (2020SYLZDXM011).

References

Effect of an Automated Writing Evaluation System on Students’ EFL Writing Performance

Limei ZHANG\textsuperscript{a} & Zheng HUANG\textsuperscript{b}
\textsuperscript{a}Nanyang Technological University, Singapore
\textsuperscript{b}Shanghai Jianqiao University, China
*limeizh2008@yahoo.com

Abstract: Recently, a growing number of studies were conducted on the use of automated writing evaluation (AWE) systems in writing classrooms. As an online writing evaluation system, Pigai has been used more and more widely in college-level English as a Foreign Language (EFL) writing courses. The current study examines the accuracy of the feedback from Pigai and the effect of the AWE system on Chinese college students’ EFL writing performance. 97 students from an academic writing course were invited to participate in the study. This study addresses the following two research questions: 1) How accurate is the feedback from Pigai in terms of the errors it can identify in students’ writing? 2) What is the effect of using the AWE system on students’ essay performance? Results showed that the average precision rate of Pigai is 58% and the accuracy varies across error types. In addition, Pigai improved students’ writing performance significantly in terms of the quality and length of essays. The findings will inform discussion of whether and how to integrate the use of online writing evaluation systems into writing classrooms.

Keywords: automated writing evaluation, Pigai, writing performance

1. Introduction

Recently, a growing number of studies were conducted on the use of automated writing evaluation (AWE) systems in writing classrooms, which includes Criterion (ETS), Project Essay Grade (PEG) Writing (Measurement Incorporated), My Access! (Vantage Learning) etc. (Attali, 2004; Chodorow, Gamon, & Tetreault, 2010; Chen & Cheng, 2008; Lavollette, Polio, & Kahng, 2015; Ranalli, Link, & Chukharev-Hudilainen, 2016). Researchers found that the use of AWE systems can affect students’ writing performance positively. For example, it was reported that the use of Criterion led to a significant decrease in the number of errors in learners’ resubmissions (Ranalli, et al., 2016).

Pigai, an AWE system developed in mainland China, has been used more and more widely in college-level English as a Foreign Language (EFL) writing courses. However, few studies have been conducted on the system itself and its effect on Chinese EFL students’ writing performance. The current study examines the accuracy of the feedback of Pigai and its effect on EFL students’ writing performance.

2. Literature Review

Automated writing evaluation (AWE) systems, which combine automated essay scoring with automated feedback (Grimes & Warschauer, 2010), provide students cycles of writing practices and formative feedback, which can reduce demands on teachers. Previous studies have shown that AWE systems can be as reliable as or more reliable than human raters in assigning scores.

An accurate AWE system can provide students with accurate information to target relevant areas of revision, improvement and learning (Attali, 2004). Consequently, they can help improve students’ writing performance (Lavollette, et. al, 2015; Ranalli, et. al, 2016). The current study examines the accuracy of the feedback from Pigai and the effect of the AWE system on Chinese college students’ EFL writing performance.
3. Research Questions

There are two research questions in the study, as follows:
1) How accurate is the feedback from Pigai in terms of the errors it can identify in students’ writing?
2) What is the effect of using the AWE system on students’ essay performance?

4. Research Methods

4.1 Participants
There were 97 first-year Chinese college students participating in this study. They were from nine different departments in a Chinese university.

4.2 The Writing Task
Participating students were required to write an essay entitled “How will AI affect our life?” The writing prompt was as follow:

Recently, Google’s AlphaGo defeated Lee Sedol, the world Go Champion, 4 to 1 in a five-game match. The machine’s sweeping victories have once again made AI (artificial intelligence) a hot topic. Some people welcome the progress and expect AI to benefit mankind in more fields. Some others fear that AI will eventually get out of control. What is your view? How will AI affect our life? Write an essay in response to the questions. Give reasons to support your points of view.

4.3 Procedure
The students were first required to complete the essay using Pigai and submit it to get the score and feedback from Pigai. Next, they were required to revise their essays following the feedback from Pigai and get a new score. Meanwhile, human raters corrected the two essays.

4.4 Data analysis
Descriptive statistics were calculated first, which includes mean and standard deviation of the scores of students’ essays. Accuracy ratings of the error types were checked. Pair-sample t-tests were conducted to compare the scores of different versions of students’ essays.

5. Results and Discussion

5.1 Descriptive statistics
Table 1 shows the descriptive statistics of the participants’ compositions.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>First draft (Pigai)</td>
<td>86.31</td>
<td>4.27</td>
</tr>
<tr>
<td>Second draft (Pigai)</td>
<td>88.96</td>
<td>3.34</td>
</tr>
<tr>
<td>Teachers’ score-1</td>
<td>10.54</td>
<td>1.38</td>
</tr>
<tr>
<td>Teachers’ score-2</td>
<td>11.28</td>
<td>1.29</td>
</tr>
<tr>
<td>Word Count-1</td>
<td>243.00</td>
<td>59.10</td>
</tr>
<tr>
<td>Word Count-2</td>
<td>248.23</td>
<td>60.59</td>
</tr>
</tbody>
</table>

5.2 Accuracy
Results showed that there were 16 types of error identified by Pigai, including four types of grammar mistakes, ten types of usage mistakes and two types of mechanic mistakes. Among the 471 mistakes identified, 272 were correct, indicating the average precision rate is 58%. For mechanic errors the precision rate is as high as 75%. Correlation between human rater’s score and Pigai’s is .59, p<.05. Table 2 shows the accuracy rating for ten most commonly identified error types.
Table 2 Accuracy ratings for ten most commonly identified error types

<table>
<thead>
<tr>
<th>Error type</th>
<th>No. of errors</th>
<th>Precision</th>
<th>Not precise</th>
<th>% Precision</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spelling mistakes</td>
<td>103</td>
<td>92</td>
<td>11</td>
<td>89.3</td>
<td>M</td>
</tr>
<tr>
<td>Wrong articles</td>
<td>17</td>
<td>12</td>
<td>5</td>
<td>70.6</td>
<td>U</td>
</tr>
<tr>
<td>Determiner noun agreement</td>
<td>30</td>
<td>21</td>
<td>9</td>
<td>70.0</td>
<td>U</td>
</tr>
<tr>
<td>Ill-formed verbs</td>
<td>21</td>
<td>14</td>
<td>7</td>
<td>66.7</td>
<td>G</td>
</tr>
<tr>
<td>Capitalization</td>
<td>43</td>
<td>26</td>
<td>17</td>
<td>60.5</td>
<td>M</td>
</tr>
<tr>
<td>Subject-verb agreement</td>
<td>102</td>
<td>53</td>
<td>49</td>
<td>52.0</td>
<td>G</td>
</tr>
<tr>
<td>Preposition error</td>
<td>8</td>
<td>4</td>
<td>4</td>
<td>50.0</td>
<td>U</td>
</tr>
<tr>
<td>Part of speech error</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>50.0</td>
<td>U</td>
</tr>
<tr>
<td>Senses error</td>
<td>10</td>
<td>4</td>
<td>6</td>
<td>40.0</td>
<td>G</td>
</tr>
<tr>
<td>Collocation error</td>
<td>32</td>
<td>9</td>
<td>23</td>
<td>28.1</td>
<td>U</td>
</tr>
</tbody>
</table>

Note: G=grammar; M=mechanics; U=usage

5.3 Effect of using AWE on students’ essay performance

Pair-sample $t$-tests were conducted with students’ first draft and second draft which were scored by Pigai. Similarly, pair-sample $t$-tests were conducted with the number of words they wrote for the two rounds of essay-writing as well as the two drafts teacher-scored essays. Results showed that $t(96)=-9.67, p<.05$; $t(95)=-7.27, p<.05$; $t(96)=-4.15, p<.05$.

As the results indicated, the feedback students received from the automated writing evaluation system, i.e., the Pigai has improved their performance significantly, in terms of the quality as well as the length of the essay. This has been supported by the results of teachers’ evaluation.

6. Conclusions and Suggestions for Future Research

The study examines the accuracy of the feedback from Pigai in terms of the errors it can identify in students’ writing and its effect on students’ writing performance. It was found that the average precision rate of the AWE system was 58%. With the use of Pigai, students’ writing performance was significantly improved.

Due to limited time and resources, the current study was tentative and exploratory in nature. It is hoped that future studies can be conducted with longer duration and larger sample sizes through multiple data collection methods.

References


Professional Learning Community’s Views on Accessibility during Emergency Remote Teaching

Tom GORHAMa & Hiroaki OGATAb

aFaculty of Letters, Rissho University, Japan
Graduate School of Informatics, Kyoto University, Japan
*gorham@ris.ac.jp

Abstract: The social distancing measures that have been adopted in response to the COVID-19 pandemic have resulted in the implementation of emergency remote teaching (ERT) practices in many educational contexts. This shift has presented several challenges, including issues of technological accessibility. A survey was conducted of members of Online Teaching Japan, a grass-roots professional learning community which was created to support teachers’ transitions to ERT. This paper will report on the results of this survey that address accessibility issues.

Keywords: Emergency Remote Teaching, Technological Accessibility, Technology Enhanced Language Learning

1. Introduction

The educational changes that have been made around the world as the result of the COVID-19 pandemic have revealed problems with equitable technological accessibility (often called a “digital divide”) around the world (UNESCO, 2020; Zhong, 2020). This paper will share some results from a survey which was conducted with members of a Japan-based professional learning community. It will report on questions that asked how concerned community members were about technological accessibility issues (e.g. a lack of reliable high-bandwidth Internet, limited Internet data plans for students relying on mobile devices, or students who do not have a desktop/laptop computer).

2. Professional Learning Community in the Age of Emergency Remote Teaching

2.1 Emergency Remote Teaching (ERT)

The COVID-19 pandemic has forced many countries to enact significant social distancing measures to prevent the further spread of the virus. Schools in Japan were affected by these measures and had to close their doors to students. Many universities in Japan announced that in addition to delaying the traditional April start of the new school year, they would also be transitioning to the delivery of their courses via the Internet.

This announcement caused distress and worry for teachers who had little to no previous experience with online teaching methodologies and technologies. Similarly, there were many schools that lacked an established plan and infrastructure for transitioning their entire student body to an online modality. Educational technology researchers have been vocal about trying to differentiate this sudden and unexpected change to online or distance education from the preferred methodical creation of a full-fledged distance education program. They stress that ERT is a temporary response to an immediate crisis and the main focus is “not to re-create a robust educational ecosystem but rather to provide temporary access to instruction and instructional supports in a manner that is quick to set up and is reliably available during an emergency or crisis” (Hodges, Moore, Lockee, Trust, & Bond, 2020, Emergency Remote Teaching, para. 1).
This surprising and stressful shift to ERT was not met with passive despair by teachers in Japan. Many came together to seek and offer support to one another in the form of Professional Learning Communities (PLCs). One such PLC is called Online Teaching Japan. The next section will offer a brief description of the group.

2.2 PLC: Online Teaching Japan

When teachers found out that they would need to transition to ERT for the start of the 2020 school year in Japan, many were anxious and wanted to learn more about how to best serve their students under the circumstances. One Tokyo-based university teacher, David Juteau, created a private Facebook group known as Online Teaching Japan on March 29, 2020. This grew into a decentralized, grass-roots Professional Learning Community for English-speaking language teachers in Japan who wanted to learn more about how to succeed while doing ERT. By early June, the group had grown to over 1,600 members. The organic growth represented a true and urgent need for professional development; it even expanded to include non-university teachers, such as secondary teachers and language school owners, and educators from outside of Japan, too.

3. Survey

3.1 Survey Materials and Description

The authors created a survey tool using Google Forms. The survey was advertised to the Online Teaching Japan PLC in three separate posts to the group’s discussion board during May 2020. This timing reflects the beginning of ERT conditions in Japan. The purpose of the survey was to gain insight into teachers’ perception and use of technology during this time. 113 people responded.

It asked five broad multiple-choice Likert scale questions. Each one was followed by an optional open-ended question which allowed the survey respondents to clarify or hedge their multiple-choice answer. All of the questions and answers were written in English.

In the interest of brevity, this survey omitted demographic questions and questions specifically about teaching context. The rationale was that the level of demographic granularity that was desired by the authors was achieved by targeting individuals who could be categorized as “people who self-selected to join the Online Teaching Japan PLC at the start of the period of ERT in Japan.”

3.2 Survey Results

This section will report on some of the qualitative and quantitative data results from the survey. It will highlight only replies to the first question (see figure 1), which asked participants about their level of concern about technological accessibility issues and its follow-up open-ended question.

Among the small percentage of respondents (7.1%) who were not concerned at all about accessibility issues, those who elected to share qualitative answers cited the privileged economic status of their students and the personal experience that their students have not reported any such problems.

On the other end of the spectrum, the respondents who were “very concerned about accessibility issues” voiced two common themes of technological and economic inequality. Many of them had personal experiences with students who have had difficulties accessing some course content and some try to remedy that problem by offering activities that have lower-bandwidth requirements:

- One of my students lives in the mountains and has no reliable internet access.
- I worry that all of the students at my school will have equal access to the education that we are trying to provide. One of the equalizing factors about schools is that students have access to equipment that they might not otherwise have access to. Not all of the students have great internet access and at .5 gb per hour of video it can eat up a data plan very quickly even if they have the extended data plan. I try my best to offer low bandwidth alternatives to my classes.
- So far, I have had minimal problems with students lacking adequate high-bandwidth Internet, but that is only because my assignments rarely require high-bandwidth Internet.
Figure 1. Survey question: How concerned are you about technological accessibility issues? (e.g. a lack of reliable high-bandwidth Internet, limited Internet data plans for students relying on mobile devices, or students who don't have a desktop/laptop computer). (n=113)

Some of the respondents who described themselves as “concerned” or “slightly concerned” about accessibility issues noted that although the majority of their students didn’t have any issues with accessibility, they did have a small number of students who did:

- I surveyed my students about this in the first week, so I know there are only a few without a computer or unlimited WiFi.
- I surveyed my students and only a few reported some issues
- Some students don’t have strong wifi
- 95% of my students have a wifi connection at home, the rest are relying on cellular data

4. Conclusion

This survey demonstrates that there is widespread concern about technological accessibility issues among members of the Online Teaching Japan PLC. Many responded that a minority of their students were experiencing some type of accessibility issues. As their institutions transition out of ERT practice, it is essential that they try to help those students overcome their accessibility issues.

Acknowledgements

We would like to thank the survey participants and the peer reviewers who helped to improve this paper.

References


Effects of Academic Achievement and Group Composition on Quality of Student-Generated Questions and Use Patterns of Online Procedural Prompts

Fu-Yun YU & Wen-Wen CHENG

* Institute of Education, National Cheng Kung University, Taiwan

wliebe@gmail.com

Abstract: The main purpose of this study was to examine if and how academic achievement and group composition affect quality of online student-generated questions. In addition, the use patterns of online procedural prompts by students of different academic achievement and gender group composition were investigated. A total of 41 intermediate-level college sophomores enrolled in an English as a foreign language class participated in this study for four weeks. An online instant interactive system, Zuvio, was adopted to support in-class student-generated questions activities. All questions students generated corresponding to the study material were categorized along the revised Bloom’s taxonomy for classifying the quality of student-generated questions, and content analysis along the integrated online procedural prompts (i.e., signal words plus the answer is; generic question stems) was adopted to reveal use patterns. Five important findings were obtained. First, students in both low- and high-academic levels generated the majority of questions in the high cognitive level. Second, more questions generated by students in all-male and mixed-gender groups fell in the high cognitive level than in the low cognitive level whereas there is an equal distribution of low and high cognitive level questions generated by the all-female group. Third, the results of the Fisher’s exact test found no significant relations between academic achievement and quality of student-generated questions. Fourth, the results of the chi-square test of independence found no significant relations between gender group compositions (i.e., all-male, all-female, mixed-gender) and quality of student-generated questions. Finally, the results of content analysis revealed that while some same use patterns of online procedural prompts were observed for students in the low- and high academic achievement levels and different gender group composition, slightly varied use patterns by students in different academic levels and gender composition were present. Suggestions for instruction and future studies are provided.

Keywords: Academic achievement, gender group composition, individual differences, online learning activity, student-generated questions, use patterns

1. Introduction

1.1 Pedagogical Values of and Support for Student-generated Questions (SGQ)

Existing studies from past decades have generally substantiated the beneficial effects of student-generated questions (SGQ) approach for enhancing understanding (Brown & Walter, 2005; Hardy, Bates, Casey, Galloway, Galloway, Kay, & McQueen, 2014; Song, 2016), learning motivation (Lam, 2014; Poot, Kleijn, Rijen, & Tartwijk, 2017), higher-order thinking (Brown & Walter, 2005; Yu & Liu, 2008), and academic performance (Hardy, et al., 2014; Khansir & Dashti, 2014; Sanchez-Elez, et al., 2014). Despite its educational benefits, quite a number of students expressed a lack of experience and confidence in SGQ tasks (Yu, 2009). In light of this, researchers proposed different pedagogical arrangements to provide support for SGQ activities.

One type of support is through the provision of procedural prompts. For instance, ‘signal words’ procedural prompt (i.e., who, what, where, when, where, and how) was suggested as one of the most frequently used and easily learned types to be introduced during SGQ for promoting students’
comprehension of learning materials (Rosenshine, Meister, & Chapman, 1996). Moreover, a set of generic question stems was proposed by Alison King (1990, 1995), including (a) questions that ask for self-generative examples, elaborated explanations or personal opinion with justifiable reasons (e.g., “How would you use ... to ...?”, “What is a new example of ...?” “Explain why...?”), “How does ... affect ...?”, and “Do you agree or disagree with this statement: ...? Support your answer.” and (b) questions directing at drawing conclusions or making differentiations and connections between prior and existing knowledge (e.g., “What conclusions can you draw about...?,” “What is the difference between ... and ...?,” and “How is ... related to ... that we studied earlier?” (p.669). The results from a series of King’s studies (1990, 1992) further showed that the devised guides positively influenced students’ elaborated responses while prompting students to generate high-order thinking questions as compared to the unguided questioning situation. In regard to ‘the answer is’ procedural prompt proposed by Stoyanova and Ellerton (1996), it has been found to support SGQ activities for math learning (Brown & Walter, 2005) and civil education learning (Yu & Pan, 2014).

Another type of support is by leveraging the power of student peers. Specifically, cooperative learning has been suggested to assist students’ inexperience in SQG tasks (Yu, Liu, & Chan, 2005) in light of its solid empirical foundations. A wealth of research has attested the efficacious effects of cooperative learning on enhancing students’ academic achievement (Gull, & Shehzad, 2015; Khan & Ahmad, 2014; Marashi & Khatami, 2017; Pan & Wu, 2013), problem-solving ability, reasoning skills (Gillies, 2011; Gillies & Haynes, 2011), creativity (Jacobs, 2017; Marashi, 2017), and motivation (Marashi & Khatami, 2017; Pan & Wu, 2013).

1.2 Factors Moderating the Effects of SGQ

As described, a set of explicit procedural prompts and pedagogical interventions have been proposed to support SQG tasks, and empirical studies are generally supportive of their respective learning effects (e.g., Yu & Pan, 2014; Yu, Tsai, & Wu, 2013). Regardless of this, some pertinent issues are still under-examined, explicitly, factors that may moderate the effects of SGQ performance — individual differences in academic performance and group composition in cooperative learning situations.

Foremost, as to individual differences in academic performance, it has been noted to affect learning process and outcomes. For instance, Efklides, Papadaki, Papantoniou, and Kiosseoglou’s empirical study (1997) found that individual ability had a direct influence on learning performance. Schmeck and Grove (1979) even provided an explanation regarding how individual differences affect learning performance — students with high achievement tended to process information comprehensively and elaborately; thus, they could retain the detailed original information better and have more organized higher-level ideas as compared to the students with low achievement.

As to group composition, studies done in the cooperative learning field have identified its effects on cooperative learning behavior and productivity. For instance, in Lee’s study (1993), it was found that students’ interactions in computer-based cooperative learning were significantly different in groups of various gender composition. When studying difference in solution-seeking behavior, Harshkamp, Ding, and Suhre (2008) revealed that female students in the mix-gender group and all-female group didn’t learn to solve physics problems and spent more time asking questions as compared to their male classmates. Zhan, Fong, Mei, and Liang’s research (2015) reported that males performed better in mixed-gender groups, but there was no difference for female performance in both same-gender groups and mixed-gender groups.

1.3 Research Questions of This Study

While empirical studies highlighted individual difference effects on learning (e.g., Efklides et al., 1997; Schmeck & Grove, 1979), currently, limited studies examine its effects in the SGQ context. Moreover, up till now, issues regarding how group composition may affect SGQ outcomes and process are yet known. Hence, in this study the researchers would like to examine whether individual differences in academic achievement and gender composition in group situations has any effects on the quality of students-generated questions and use patterns of the integrated procedural prompts.

In specific, four research questions are proposed:
RQ#1: Whether academic achievement has any relations to the quality of SGQ?
RQ#2: Whether gender group composition has any relations to the quality of SGQ?
RQ#3: Whether academic achievement has any relations to the use pattern of online procedural prompts?
RQ#4: Whether gender group composition has any relations to the use pattern of online procedural prompts?

2. Methods

2.1 Participants and The Study Context

Forty-one intermediate-level college sophomores (males: 22) enrolled in an English as a foreign language class from the College of Management at a National University in southern Taiwan were invited to participate in this study, which lasted for four weeks. To promote student learning of the learning material, the researchers implemented the SQG approach in two sessions of this class. In particular, Stoyanova and Ellerton’s (1996) ‘the answer is’ with ‘signal words’ and King’s ‘generic question stems’ were selected as procedural prompts for SQG. An online instant interactive system, Zuvio, was introduced to support SGQ (see Figure 1). The participants could access Zuvio by any portable device of their choice (e.g., smartphones, laptop, tablets, etc.) to generate and submit questions with answers on the learned content.

For the purpose of this study, one unit with 4 lessons on the topic of Inventions and Discoveries was selected. Each lesson focused on the photo story (i.e., the topic), vocabulary (on technology), grammar (on past unreal conditional), and an article (on antibiotics), respectively. A brief training session on SGQ in Zuvio was arranged before the 1st SGQ activity to ensure that the participants were equipped with associated knowledge and skills for meaningful engagement in the online activity.

Two online SQG activities were scheduled after the mid-term exam. The ‘signal words plus the answer is’ procedural prompts were used for the 1st SQG activity after the 2nd lesson, during which each student generated one question corresponding to the delivered instruction individually (see Figure 1). For the 2nd SQG activity, it was scheduled after the 4th lesson, and the ‘question stems’ procedural prompt was introduced, during which students generated two questions in correspondence to the taught instruction in groups of two (see Figure 2).

![Instruction and Procedural Prompts Provided for the 1st SGQ Activity](image)

*Figure 1. Instruction and Procedural Prompts in the Form of ‘The Answer is’ and ‘Signal Words’ Provided for the 1st SGQ Activity on Zuvio*
2.2 Classification of the Quality of SGQ

In total, 123 questions were generated by the participating students in the two online SGQ activities. The revised Bloom’s Taxonomy (Anderson & Krathwohl, 2000), which has been widely used for evaluating the cognitive levels of questions in textbooks (Assaly & Smadi, 2015; Tarman & Kuran, 2015) and assessing SGQ performance (Lameese, Madalyn, Keli, Matthew, Jakob, Christina, 2015) was adopted and operationalized for classifying the quality of SGQ (see Table 1).

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remember</td>
<td>Q&amp;A* involves recalling information in the textbook.</td>
</tr>
<tr>
<td>Understand</td>
<td>Q&amp;A* involves describing information in the textbook.</td>
</tr>
<tr>
<td>Apply</td>
<td>Q&amp;A* involves using information in new situations.</td>
</tr>
<tr>
<td>Analyze</td>
<td>Q&amp;A* involves identifying cause-and-effect or analyzing a problem.</td>
</tr>
<tr>
<td>Evaluate</td>
<td>Q&amp;A* involves making judgments about the focal content.</td>
</tr>
<tr>
<td>Create</td>
<td>Q&amp;A* involves synthesizing multiple units of information into new coherent entity or providing new solution.</td>
</tr>
</tbody>
</table>

*A question with its answer

Two experienced English teachers independently categorized each of the 123 questions the participants generated during the two online SGQ sessions along the revised Bloom’s taxonomy. Percent of agreement was adopted for inter-rater reliability and evidenced adequate reliability — 82.96% and 84.38% for the 1st and 2nd SGQ activities, respectively.

2.3 Data Analysis of SGQ

Besides descriptive statistics (i.e., frequency, percentage), the Pearson’s chi-square test was adopted to analyze whether academic achievement and gender group composition, respectively, has significant association with the quality of SGQ. In view of the fact that 33.33% of the cells in the contingency table had a number less than 5, to ensure valid chi-square tests and to comply with the calculation rule (i.e., requiring at least 80% of the cells to have an expected count greater than 5), the cognitive levels were
grouped into a low level (by combining the bottom three cognitive levels: remember, understand, and apply) and a high level (by combining the top three cognitive levels: analyze, evaluate, and create).

For classifying students’ academic achievement levels, it was originally based on the Common European Framework of Reference for Languages (CEFR) and language standard of English proficiency tests issued by Ministry of Education (2016) and TOEIC scores, and grouped to three levels (i.e., below 350 points as the low level, 351 to 550 points as the medium level, and above 551 as the high level). To comply with the chi-square calculation rule while considering approximately equal number in different groups, students’ academic achievement levels were grouped to two levels, with below 450 points as the low-achieving level and above 451 points as the high-achieving level.

Finally, content analysis was applied for examining the use patterns of the integrated procedural prompts.

3. Results

3.1 RQ#1: Relations of Academic Achievement and Quality of SGQ

In total, 41 questions were generated during the 1st online SGQ activity. As shown in Table 2, the majority of SGQ from both the low- and high-achieving students were at the high cognitive level. Question at the high cognitive level generated by the high- and low-achieving students include content not based solely on the textbook, but with reference to other sources (e.g., personal experiences, daily life, internet). Sample questions include: How does the robot vacuum work? Is it efficient? (A: Yes. I don't waste time on sweeping the floor after I bought it; target word: efficient); Why is that car so expensive? (A: Because it’s not only a Ferrari, but also a unique model; target word: unique).

Fisher’s exact test considered a better method was adopted here instead because the data dealt with a 2x2 contingency table with small sample sizes, and there were two observed values less than 5. The results showed that there were no significant relations ($p = 1.000 > 0.05$) between students’ academic achievement and quality of SGQ.

<table>
<thead>
<tr>
<th>SGQ Cognitive Levels</th>
<th>Low (%)</th>
<th>High (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low achieving</td>
<td>4 (16%)</td>
<td>21 (84%)</td>
</tr>
<tr>
<td>High achieving</td>
<td>2 (12.5%)</td>
<td>14 (87.5%)</td>
</tr>
<tr>
<td>Total</td>
<td>6 (14.6%)</td>
<td>35 (85.4%)</td>
</tr>
</tbody>
</table>

3.2 RQ#2: Relations of Gender Group Composition and Quality of SGQ

In total, 82 questions were generated during the 2nd online SGQ activity. As shown in Table 3, in the all-male and mixed-gender groups, more questions generated fell in the high cognitive level than in the low cognitive level whereas for the all-female group, there was an equal distribution of questions falling in low and high cognitive levels.

The results of chi-square test of independence further showed that there were no significant relations ($p = 0.443 > 0.05$) between gender group composition and quality of SGQ.

<table>
<thead>
<tr>
<th>SGQ Cognitive Levels</th>
<th>Low (%)</th>
<th>High (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All-male</td>
<td>14 (38.9%)</td>
<td>22 (61.1%)</td>
</tr>
<tr>
<td>All-female</td>
<td>12 (50%)</td>
<td>12 (50%)</td>
</tr>
<tr>
<td>Mixed-gender</td>
<td>7 (31.8%)</td>
<td>15 (68.2%)</td>
</tr>
<tr>
<td>Total</td>
<td>33 (40.2%)</td>
<td>49 (59.8%)</td>
</tr>
</tbody>
</table>
3.3 RQ#3: Relations of Academic Achievement and Use Pattern of Online Procedural Prompts

The questions generated during the 1st online SQG activity were further analyzed along the ‘signal words’ procedural prompt to examine their respective uses by the low and high-achieving participating students. As shown in Table 4, both ‘what’ and ‘why’ signal words were used for SQG by both low and high-achieving students, with ‘why’ being used most frequently, followed by ‘what’ by both high- and low-achieving students. Moreover, the ‘when’ signal word was never used by neither group. Despite the two same use patterns by high- and low-achieving students, some different use patterns were present. Explicitly, ‘who’ was used exclusively by low-achieving students whereas ‘how’ and ‘where’ were used only by high-achieving students.

Table 4. Use of Signal Words by the Low- and High-Achieving Students

<table>
<thead>
<tr>
<th>Signal Words</th>
<th>Low achieving f(%)</th>
<th>Where</th>
<th>When</th>
<th>Why f(%)</th>
<th>Who f(%)</th>
<th>How</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low achieving</td>
<td>7 (27%)</td>
<td>0</td>
<td>0</td>
<td>14 (53.8%)</td>
<td>5 (19.2%)</td>
<td>0</td>
</tr>
<tr>
<td>High achieving</td>
<td>3 (20%)</td>
<td>2 (13.3%)</td>
<td>0</td>
<td>9 (60%)</td>
<td>0</td>
<td>1 (6.7%)</td>
</tr>
<tr>
<td>Total</td>
<td>10 (24.4%)</td>
<td>2 (4.9%)</td>
<td>0</td>
<td>23 (56.1%)</td>
<td>5 (12.2%)</td>
<td>1 (2.4%)</td>
</tr>
</tbody>
</table>

3.4 RQ#4: Relations of Gender Group Composition and Use Pattern of Online Procedural Prompts

The questions generated during the 2nd online SQG activity were further analyzed along the set of online ‘generic question stems’ procedural prompt to examine their respective uses by the participants of different gender group composition. As shown in Table 5 (the right-most column), as a whole, among the 13 question stems, only about half (i.e., 7 question stems) were used for the SQG activity. Moreover, among the seven used question stems, three question stems were used by all three groups, leading to the most frequently used question stems — ‘What is the difference between … and …?’; ‘Explain why …?’, and ‘What do you think would happen if …?’ in that order.

Table 5. Use of ‘Generic Question Stems’ by Different Group Composition

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. How would you use … to…?</td>
<td>2</td>
<td>4</td>
<td></td>
<td>6 (7.3%)</td>
</tr>
<tr>
<td>2. What is a new example of …?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. What is the difference between … and …?</td>
<td>14 [1]</td>
<td>4</td>
<td>2</td>
<td>20 (24.4%) [1]</td>
</tr>
<tr>
<td>6. How are … and … similar?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. What is a possible solution to the problem of …?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. What conclusions can you draw about …?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. How does … affect…?</td>
<td>2</td>
<td>5</td>
<td></td>
<td>7 (8.5%)</td>
</tr>
<tr>
<td>10. In your opinions, which is best, … or …?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Why?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. What are the strengths and weaknesses of …?</td>
<td>7 [2]</td>
<td>3</td>
<td></td>
<td>10 (12.2%)</td>
</tr>
<tr>
<td>12. Do you agree or disagree with this statement…? support your answer.</td>
<td>4</td>
<td></td>
<td></td>
<td>4 (4.9%)</td>
</tr>
<tr>
<td>13. How is … related to … that we studied earlier?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>36</td>
<td>24</td>
<td>22</td>
<td>82 (100%)</td>
</tr>
</tbody>
</table>
As for the use pattern differences among the three different group composition, more question stems were used by the all-male group (i.e., seven) as compared to the other two groups (i.e., 5 question stems used by the all-female group and 4 by the mixed-gender group). Moreover, ‘What is the difference between … and …?’ was the question stem used most by the all-male group whereas ‘Explain why …?’ was used most by both all-female and mixed-gender groups.

4. Discussion and Conclusion

In view of the existing literature on educational psychology that points to possible effects of individual differences (e.g., Efklides et al., 1997; Schmeck & Grove, 1979) and group composition under a cooperative learning situation (e.g., Harskamp et al., 2008; Lee, 1993; Zhan et al., 2015), issues regarding if and how such factors may have on SGQ was the focus of this study. Specifically, individual differences in academic achievement and gender composition were targeted, and their respective relations to the quality of SGQ as well as use pattern of integrated online procedural prompts were examined.

The results on Fisher’s exact test showed that there were no significant relations between students’ academic achievement and the quality of SGQ (in terms of cognitive level). Although this study did not concur with past studies confirming individual differences effects, with students in both low- and high-academic levels generating the majority of questions in the high cognitive level, the authors speculated that it may be the explicit nature of procedural prompts (i.e., ‘signal words plus the answer is’) that help guide the participating students in generating high-level cognitive level questions; thus, it helps alleviate English capability gap between students at different English achievement levels.

Furthermore, the results of chi-square test of independence found no significant association between gender group composition and cognitive levels of SGQ. Again, despite that this study did not corroborate with existing studies attesting gender effects in cooperative group situations, with more questions generated by students in the all-male and mixed-gender groups falling in the high cognitive level than in the low cognitive level and an equal distribution of questions in both low and high cognitive levels by the all-female group, it appeared that the set of online procedural prompts provided is successful in directing the participants in different gender compositions not to delimit question-generation in the low-cognitive end, as so concerned by practitioners (King, 1990, 1992).

Lastly, while some same use patterns of online procedural prompts were observed for students in the low- and high academic achievement levels (e.g., use of the ‘why’ signal word most frequently, followed by the ‘what’ signal word) and different gender group composition (e.g., nearly half of the provided prompts were not used), slightly different use patterns were present for students in different academic achievement levels and gender composition.

4.1. Limitations and Suggestions for Instructors and Future Studies

The current study found that under the online provision of ‘signal words plus the answer is’ and ‘question stems’ procedural prompts, students with different academic performance levels and gender group compositions were found to generate significantly more questions at the high cognitive level. With a considerable proportion of students lacking experience in SGQ and worrying about their performance at the SGQ task (Yu, 2009), it is suggested that instructors take advantage of the explicit nature of procedural prompts to support online SGQ for high cognitive level question-generation.

This empirical study provided preliminary data on the relations of academic performance and gender group composition on the quality of SGQ and use patterns of online procedural prompts. Nonetheless, it should be noted that only one class of undergraduate students was involved to participate in two online SGQ activities in corresponding to a study unit on English. For future studies, larger sample sizes for an extended study period involving different topics should be considered. With an extended study period across different topics, the use pattern of different procedural prompts across different topics can be better examined and understood. With larger sample sizes, the learning processes and outcomes of male and female students in different gender compositions can be examined and compared to the findings of previous studies for better understanding (e.g., the study results of Zhan et
al. found that males performed better in mixed-gender groups while their female counterparts performed similarly in both same-gender groups and mixed-gender groups). Moreover, future studies incorporating qualitative research method, in particular, in-depth interview, would be able to probe deeper and gain insight as to why and how different online procedural prompts are considered by students in different academic achievement levels and gender group composition for SGQ. Finally, individual differences in other aspects found to affect learning, for instance, general cognitive abilities and functioning (Gustafsson & Undheim, 1996; Ruiz, Chen, Rebuschat, & Meurers, 2019) may be better tapped with a larger pool of participants.

References


Assaly, I. R., & Smadi, O. M. (2015). Using Bloom’s taxonomy to evaluate the cognitive levels of master class textbook’s questions. English Language Teaching, 8(5), 100-110.


School and teacher level predictors for students’ information literacy in Chinese rural and urban education

Liqin Yu a, Di Wu b*, Sha Zhu a, Feixiong Chen a, Kaige Zhu c

a National Engineering Research Center for E-Learning, Central China Normal University, Wuhan, China
b Educational Informatization Strategy Research Base Ministry of Education, Central China Normal University, Wuhan, China
c College of Liberal Arts, Ping Ding Shan University, Ping Ding Shan, China
*wudi@mail.ccnu.edu.cn

Abstract: This study aimed to investigate the level of secondary school students’ information literacy in China and examine the contribution of school and teacher level factors on students’ information literacy between rural and urban schools. A total of 598 schools, 56415 students, and 18286 teachers participated in this study. The findings of this study were as follows: (1) the overall of secondary schools students’ information literacy only reached an average level and urban school students’ information literacy were significantly higher than that of rural school students; (2) In rural schools, teachers’ ICT collaboration was a positive predictor for students’ information literacy, while teachers’ ICT use for learning was identified as a negative predictor of students’ information literacy; (3) In urban schools, ICT management, ICT operation, and teachers’ ICT self-efficacy were found to be significantly associated with students’ information literacy. Based on the findings, suggestions for improving students’ information literacy between rural and urban schools were discussed.

Keywords: Information literacy, Chinese secondary school students, rural and urban school.

1. Introduction

Information literacy is gaining recognition as being vital for students living, learning and working in the twenty-first century (European Commission, 2007; Kim & Lee, 2013) and the school has been widely acknowledged as an important place for equipping their students with new kinds of skills such as information literacy (Gerick, 2018). Previous studies have revealed that school and teacher level factors such as ICT infrastructure, ICT classes, teachers’ ICT attitude, ICT self-efficacy, ICT use in class and ICT collaboration were major predictors for students’ information literacy (Kim et al., 2011; Zhong, 2011; Zhu et al., 2019; Guo & Tsai, 2014).

In recent years, the Chinese government has announced a series of education policies to support the cultivation of students’ information literacy. For instance, in 2017, A new information technology curriculum standard was published for high schools to assist students to use computers with greater fluency (The ministry of education, 2017). To accurately measure students’ information literacy level, the Chinese ministry of education regards the assessment of students’ information literacy as an important task for the development of education information in the 2.0 age (The Ministry of Education, 2018).

Despite these efforts, there are still several problems that exist in Chinese information literacy education. A study showed that the overall level of Chinese secondary students was just “pass” (Li & Ranieri, 2010). Some researchers also pointed out that the development of students’ information literacy was unbalanced in China (Zhang & Zhu, 2016). What is the level of students’ information literacy in urban and rural schools in China? Is a significant influence of school and teacher level factors on students’ information literacy between rural and urban schools? To best of our knowledge, little research has addressed these issues, especially in developing countries such as China.
Therefore, this study aims to investigate the students’ information literacy concerning rural and urban schools from a national wide perspective. What’s more, this study also aims to analyze the schools and teacher level predictors for students’ information literacy in comparison to rural and urban schools. The results of our study are expected to help policymakers and authorities to understand the status of students’ information literacy in China and identify effective strategies and policies to reduce the digital divide of students’ information literacy between rural schools and urban schools.

2. Literature review

2.1 The definition of information literacy

Since the term of information literacy was coined by in 1974 by Paul Zurkowski, the concept of information literacy has been influenced by the evolution of the information society. Various institutions and researchers have put forward different definitions of information literacy. For example, UNESCO (2003) defined information literacy as the ability to identify, locate, evaluate, organize, and effectively create, use, and communicate information to address. Another resembling definition was provided by the Educational Testing Service (ETS), which treated the information literacy as the ability to use digital technology and networks to access, manage, integrate, evaluate, and create information (ETS, 2007). Besides, The Association of College and Research Libraries (ACRL) made a definition of information literacy as a set of comprehensive abilities encompassing the reflective discovery of information, the understanding of how information is produced, and the recreating new knowledge (ACRL, 2016). Furthermore, the Chartered Institute of Library Information Professionals (CILIP) put forward a new definition of information literacy, which emphasized the ability to think critically and express informed views (CILIP, 2018).

More recently, with the rapid development of art intelligence, big data and cloud computing, computational thinking (CT) is becoming an important element for future talents and many institutions have taken CT as a new dimension of students’ information literacy (Bae et al. 2017; Kim, Ahn, & Kim, 2019; IEA, 2016). By consolidating the existing definition of information literacy and based on our previous studies (Zhu et al. 2017; Zhu, Yang, MacLeod, Yu, & Wu, 2019), four dimensions of information literacy in this study were been proposed as following: Information Awareness and Attitude, Information Knowledge and Skills, Information Thinking and Behavior, and Information Social Responsibility. Awareness and Attitude refer to one’s information sensitivity including perception awareness, application awareness, and security awareness. Knowledge and Skills include fundamental knowledge of network, internet, PC, and a set of skills involved in the ability to use ICT. Thinking and Behavior involve the ability to think critically and use the appropriate information technology to solve complex problems, create and express ideas compellingly. Social Responsibility refers to moral principles and understanding of the rules governing information activities.

2.2 School factors influencing students’ information literacy

Concerning school-level variables, prior studies have reported that ICT infrastructure, school size, computer curriculum, and ICT management were major predictors for students’ information literacy. For example, some studies revealed that ICT availability at schools and the proportion of ICT equipment per student own were significantly associated with students’ information literacy (Kim, Kil & Shin, 2014; Zhong,2011; Seo et al., 2009). Kim (2014) found that students who had a higher completion rate of computer-related coursed showed a relatively high level of information literacy. Similarly, other studies showed the number of ICT classes positively correlated with the grade of students’ information literacy (Baek et al., 2008; Kim et al., 2011). As for the impact of school location on students’ information literacy, the results were inconsistent. For instance, Kim et al. (2011) reported that students living in urban areas have a higher ICT literacy level than do students living in rural areas. Whereas in other studies, students who live in provincial areas had superior information literacy compared with students living in major cities (Seo et al., 2009; Baek et al., 2008). However, little research has examined the schools’ influence on secondary students’ information literacy between rural schools and urban schools in China.
2.3 Teacher factors influencing students’ information literacy

Regarding teacher-related variables, precedent studies reported that teachers’ ICT capabilities, ICT attitude, ICT self-efficacy, and ICT usage were major influential factors of students’ information literacy. For example, Meelissen and Drent (2008) claimed that teachers’ attitudes towards ICT had an indirect effect on students’ information literacy through influencing students’ ICT attitude. Aesaert, Vanderlinde, and Tondeur (2015) reported that ICT usage in class was associated with students’ information literacy. Teachers’ ICT self-efficacy refers to their belief in completing ICT-related tasks. Previous studies found that teachers’ ICT self-efficacy was a positive determinant of students’ information literacy (Meelissen & Drent, 2008; Zhu et al, 2019). For the teachers’ ICT collaboration, Lai, Guo, and Tsai (2014) claimed that a collaborative teaching approach had a positive impact on students’ information literacy. As far as we know, no study has investigated the impact of teachers’ factors on secondary students’ information literacy between rural and urban schools in China.

2.4 The present study

To balance the development of students’ information literacy in China, it is necessary to understand the status of students’ information literacy and analyze the key predictors of students’ information literacy concerning rural and urban schools. Although there is an extensive body of studies that have documented several influential factors of students’ information literacy, little research has analyzed predictors of students’ information literacy by differentiating the type of schools. Besides, no large-scale assessment has been conducted so far to investigate the students’ information literacy in developing countries such as China. Therefore, this study aims to assess Chinese secondary schools’ students’ information literacy skills and examine the different influential factors of students’ information literacy by comparison between rural and urban schools. The following research questions are addressed in this study:

RQ1: What’s the level of students’ information literacy between rural schools and urban schools?
RQ2: What are the major predictors of students’ information literacy at the school level? Is there a difference between urban school and rural schools?
RQ3: What are the major predictors of students’ information literacy at the teacher level? Is there a difference between urban school and rural schools?

3. Methodology

3.1 Sampling

This study was conducted from October 2018 to December 2018. Three-stage of sampling method were used to collect data. In the first stage, 368 municipal and county areas were selected from 31 provinces in China according to their economic level. Well-developed and underdeveloped areas are in half in each province respectively. In the second stage, 3 to 5 junior schools were selected within each of the selected areas. The proportion of rural schools and urban schools was equal. In the third stage, students from seventh and eighth grades were randomly selected from each school. Ninth-grade students were excluded from this survey because of entrance academic stress. This survey included 598 schools, 64.05% of them were urban schools and 35.95% of them were rural schools. A total of 56415 students and 18286 teachers also participated in this survey. Among these participants, 36162 students and 11712 teachers were from urban schools, and 20253 students and 6574 teachers were from rural schools.
3.2 Instrumentation

The instruments of this study included three parts:
1) Students’ information literacy test. A total of 41 multiple choice questions on the web platform were designed to measure students’ information literacy. The four dimensions of information literacy are as following: information awareness and attitude (10 items), information knowledge and skills (15 items), information thinking and behavior (10 items), and information social responsibility (6 items). The overall reliability coefficients (a) of students’ information literacy test was 0.84. Information awareness and attitude refer to students’ sensitivity and judgment on information, including information perception awareness, application awareness, and security awareness. A sample item is, “What should you do if you travel to a strange place and get lost”. Information knowledge and skills refer to students’ information science knowledge and the skills to use specific ICT applications such as word, excel and Photoshop, etc. A sample item is, “Which of the following devices belongs to the computer output device?”. Information thinking and behavior refer to students’ ability for using technological tools to take part in learning activities independently and innovatively. For example, students are asked to draw the structure of a learning topic using a mind mapping tool. Information social responsibility refers to moral principles and understanding of the rules governing information activities. A sample item is, “The following ACTS in accordance with the network ethics is?”. 

2) A school questionnaire. The school questionnaire consisted of 21 web-based items to collect data about school ICT infrastructure, ICT resources, ICT operations, teacher ICT tanning, and ICT management. The items of the school questionnaire were adopted from a previous study (Wu, Li, Zhou, Tsai, & Lu, 2019). The Cronbach’s alpha values for these scales were 0.70, 0.65, 0.68, 0.72, and 0.60 respectively.

3) A teacher questionnaire. Teachers’ data were collected via five scales including ICT self-efficacy (14 items), ICT use for teaching (10 items), ICT use for students’ learning (13 items), ICT collaboration (5 items), and ICT attitude (8 items). All the items of the teacher questionnaire were adopted from previous studies (Luan, Fung, Nawawi, & Hong, 2005; Aesaert et al., 2015). The Cronbach’s alpha values obtained in this study were as follows: ICT self-efficacy (alpha=0.92), ICT use for teaching (alpha= 0.92), ICT use for students’ learning (alpha=0.97), ICT collaboration (alpha= 0.86), and ICT attitude (alpha= 0.86).

3.3 Data collection and analysis Procedures

With the help of provincial education administrative departments and local education administrative departments, students were arranged in the computer lab of each sample school to complete the information literacy test. At the same time, teachers and chief of educational information of each selected school were required to finish the teacher questionnaire and school questionnaire respectively. Students and teachers were matched through the schools’ names. All participants were informed of the research purposes and were required to sign formal consent to participate in the study. SPSS 22.0 software was used in this study. Descriptive statistics were used to describe the overall level of students’ information literacy and regression analyses were conducted to explore the effect of school and teacher level factors on students’ information literacy in rural and urban schools respectively.

4. Results

4.1 Students’ information literacy between urban and rural schools

To answer Q1, t-test companions were conducted to examine the difference in students’ information literacy between urban and rural schools. On average, the information literacy of students from urban schools is 60.99, and the information literacy of students from rural schools is 54.94. The information literacy and other dimensions including information awareness and attitude, information knowledge
and skills, information thinking and behavior, information social responsibility of students from urban schools were significantly better than that of students from rural schools.

Table 1. Students’ Information Literacy Comparison Between Urban and Rural Schools

<table>
<thead>
<tr>
<th></th>
<th>Urban</th>
<th>Rural</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Information literacy</td>
<td>60.99</td>
<td>15.55</td>
<td>54.94</td>
</tr>
<tr>
<td>Awareness and cognition</td>
<td>28.82</td>
<td>7.40</td>
<td>26.55</td>
</tr>
<tr>
<td>Knowledge and skills</td>
<td>9.88</td>
<td>3.12</td>
<td>8.74</td>
</tr>
<tr>
<td>Thinking and behavior</td>
<td>9.58</td>
<td>2.53</td>
<td>8.48</td>
</tr>
<tr>
<td>Social responsibility</td>
<td>12.71</td>
<td>4.40</td>
<td>11.18</td>
</tr>
</tbody>
</table>

*Note: **p<0.01, ***p<0.001*

4.2 School-level factors predicting students’ information literacy

To answer the second question, stepwise regression analysis was conducted to explore the relationship between school-level factors and students’ information literacy, as shown in Table 2. School ICT related factors were viewed as predictors to explain the variations in students’ information literacy. In urban areas, ICT operations (t=2.21, p<0.05) and ICT management (t=2.40, p<0.05) could make significant predictions (6 % explained) for the students’ information literacy. While in rural areas, school-level factors were found less significantly associated with students’ information literacy.

Table 2. The Regression Analysis of School-level Factors

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Rural</th>
<th>Urban</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE</td>
</tr>
<tr>
<td>ICT infrastructure</td>
<td>1.98</td>
<td>7.58</td>
</tr>
<tr>
<td>ICT resource</td>
<td>8.95</td>
<td>8.13</td>
</tr>
<tr>
<td>ICT operations</td>
<td>5.19</td>
<td>6.80</td>
</tr>
<tr>
<td>Teacher ICT training</td>
<td>-4.91</td>
<td>6.70</td>
</tr>
<tr>
<td>ICT Management</td>
<td>0.13</td>
<td>8.06</td>
</tr>
</tbody>
</table>

4.3 Teacher-level predictors on students’ information literacy

To answer RQ3, stepwise regression was employed to investigate the relationship between teacher-level factors and students’ information literacy. As shown in Table 3. In rural schools, ICT collaboration could make a positive significant prediction for students’ information literacy. However, teachers’ ICT use for learning made a negative influence on students’ information literacy. In urban schools, only teachers’ ICT self-efficacy did a significant impact on students’ information literacy.

Table 3. The Regression Analysis of Teacher-level Factors

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Rural</th>
<th>Urban</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE</td>
</tr>
<tr>
<td>ICT self-efficacy</td>
<td>21.09</td>
<td>20.92</td>
</tr>
<tr>
<td>ICT attitude</td>
<td>-22.00</td>
<td>20.84</td>
</tr>
<tr>
<td>ICT collaboration</td>
<td>54.43</td>
<td>19.73</td>
</tr>
<tr>
<td>ICT use for teaching</td>
<td>42.28</td>
<td>23.13</td>
</tr>
<tr>
<td>ICT use for learning</td>
<td>-62.56</td>
<td>25.24</td>
</tr>
</tbody>
</table>
5. Discussion and conclusion

The results of this study demonstrated that the overall Chinese secondary school students’ information literacy only reached an average level. However, it must be noted that the significant diversity of students’ information literacy still exists in rural and urban schools. The Chinese government should pay more special efforts to reduce the digital divide in terms of students’ information literacy (Chetty et al., 2018). More importantly, this study analyzed different factors affecting students’ information literacy from school and teacher aspects in rural and urban schools. The results could provide more insight for understanding differentiated needs regarding information literacy education between rural and urban schools.

In rural schools, the results indicated that teachers’ ICT collaboration significantly associated with students’ information literacy, the result was consistent with the previous studies (Lai, Guo & Tsai, 2014; Zhu et al. 2019). Teachers were found to feel less work stressful, gain a better understanding of the curriculum and be more willing to adopt new technology in the classroom through mutual collaboration among colleagues (Eickelmann, 2010; Fraillon et al., 2014; Cheung & Slavin, 2012). A collaborative atmosphere of school culture and regular ICT training activities can contribute to ICT-related collaboration among teachers (Drossel, Eickelmann, & Schulz-Zander, 2017). However, it is should be noted that rural teachers’ ICT use for learning was found to be a negative predictor of students’ information literacy. This result could be explained that due to the lack of ICT integration knowledge and ICT related competence, rural school teachers often improperly selected ICT tools in the teaching activities without the aim of developing students’ information literacy (He & Wray, 2017).

In urban schools, the results indicated that school-level factors such as ICT management and ICT operation were positive predictors of students’ information literacy. The findings are in line with earlier studies, which claimed that ICT supporting conditions is a major challenge facilitating ICT application in schools and school leadership can be identified as relevant for students’ acquisition of information literacy (Wu et al., 2019; Lorenz, Eickelmann, & Gerick, 2015). As for teacher-level factors, only teachers’ ICT self-efficacy was found to be positively associated with students’ information literacy. This result implied that urban schools’ teachers were more confident to use ICT in daily instruction which had a positive impact on students’ information literacy (Aesaert et al., 2015; Papastergiou, 2010). To conclude, secondary school students’ level of information literacy has much room for improvement in China and there is a significant difference in influencing factors of students’ information literacy between urban and rural schools in terms of school and teacher level factors. Differentiated strategies are needed to be considered for improving students’ information literacy between rural and urban schools. Briefly, the results of this study also confirm that the digital divide represents a big social challenge on a global level and reveals that schools and policy-makers still have to develop effective strategies to improve students’ information literacy both in China and in Western countries.

Acknowledgements

This work was supported by The Chinese ministry of education's major research on philosophy and social sciences project “Research on Internet + education system” (No. 16JZD043).

References


The Ministry of education (2017), Information technology curriculum standards for high schools Retrieved July 12,2019 from http://www.szsjyw.com/dede/uploads/soft/180116/%C6%D5%CD%A8%B8%DF%D6%D0%D0%C5%CF%A2%BC%BC%CA%F5%BF%CE%B3%CC%B1%EA%D7%BC2017%B0%E6.pdf.


Levels of academic teachers digital competence: Polish case-study

Nataliia DEMESHKANT*, Katarzyna POTYRAŁA* & Łukasz TOMCZYK
*Pedagogical University of Cracow, Poland
*nataliia.demeshkant@up.krakow.pl

Abstract: The purpose of this study is to examine levels of academic teachers’ digital competences regarding the demographics and professional backgrounds in Polish educational context. The 2-factor (Pedagogical and Technological knowledge) TPACK model is used in the study. A survey was administered to 103 academic teachers from Polish universities. Descriptive analysis indicated a significant negative correlation between some demographic variables (age, years of teaching, titles and degrees) and domains from both factors. However, some positive associations with certain variables from the professional background (using for teaching online learning environments, digital quizzes or polls, interactive apps or games; providing on-line courses; creating videos for teaching) were also noticed. Implications for professional development and suggestions regarding teachers’ digital competences and TPACK have been discussed.

Keywords: academic teachers; digital competence; TPACK; demographics and professional backgrounds

1. Introduction

Professional teachers’ digital competences are seen as a combination of professional, pedagogical and technological knowledge and skills (Koehler & Mishra, 2009), but also the ability to apply learning outcomes as is appropriate to the context (CEDEFOP, 2014). Most educators have been led to believe that the integration of technology into classrooms, or the transition from face-to-face learning to online learning, is a simple task requiring merely some technology skills training (Schmid & Hegelheimer, 2014). Professional development for educators moving to an online environment frequently provides only instruction on how to use the new Learning Management System and other technical skills, without a discussion about the content or pedagogical issues that intertwine with technology (Benson & Ward, 2013). The ability to use a variety of technologies did not necessarily result in the effective use of technology to impact teaching or learning. Technology skills learned in isolation may even hurt an instructor’s ability to see the complex application of that technology in a pedagogically and contextually sound manner (Benson & Ward, 2013). Therefore, the role of teachers and the systems around them must be reinvented. Many countries are currently in the process of developing or revising frameworks and training programs to guide teacher training and continuous professional development in this area (Castéra et al., 2020; Cubeles et al., 2018; Ghomi & Redeker, 2019; Starkey, 2020).

During the last few years, Polish universities also have been making conscientious efforts to improve the educational process with modern digital teaching and learning methodology (The digitalisation of Polish Education Vision and proposals, 2016). According to the Digital Economy and Society Index (DESI) profile developed by the European Commission, Poland is in the group of countries with low level of digitization (European Commission, 2019). Significant changes connected with informatization in the Polish higher education system have been done in the sphere of university management, reporting, scientific and research activities as well as teaching digitalization. Despite this, there is no clear proof about using theoretical scientific conceptions for academic teachers’ development.

The main research question was to explore how demographics and professional backgrounds of academic teachers are correlated with their levels of digital competences. Results of the study would be useful for selecting effective methods of teachers’ professional development for promoting ICT in education.
2. Theoretical background

The study is focused on academic teachers’ TPACK as one of the most important elements in teachers' professional training and development. TPACK adds technological knowledge as a new component that has to blend in with domain and pedagogical knowledge to effectively integrate ICT in instructional practices (Voogt & Mckenney, 2016). As the results of the previous studies indicate, there is no scale using the TPACK framework, which is suitable for all settings – in-service, pre-service, academic teachers, different subjects, and different countries (Cubeles et al., 2018). Controversial results have also been found in terms of how demographic data correlates with TPACK components and which TPACK components are rated higher. One more reason for our study is that the TPACK model has been applied principally in primary and secondary education and its use in the university sector is still in its initial phases and the role of the university professor has yet to be fully defined (Cubeles et al., 2018). There is a significant difference between the functions of school teachers and academic lectures. Academic teachers combine research activity with the transfer of knowledge to students. For modern academic teachers, the technological side of ICT is a part of their routine professional work, including teaching. Nevertheless, there are doubts about the pedagogical context of effective and methodically correct use of ICT for work with students. For this reason in our study, we decided to modify the classic 7-factor model structures of the TPACK framework and to combine some original factors. In this way PK, CK, and PCK were merged into one Pedagogical Knowledge factor F1 (PK). All items with technology (TK, TCK, TPK, and TPACK) were merged into second Technological Knowledge factor F2 (TK). TK factor meaning that academic teachers perceive technology already integrated with content and pedagogy. The use of the 2-factor TPACK model allows us to thoroughly examine the basic components (TK and PK) of the study’s subject (digital competences of academic teachers), as well as to determine more precisely the relationships regarding this components’ mutual dependence and development.

Besides TPACK theory there is The European Framework for the Digital Competence of Educators (DigCompEdu) describes the digital competences specific to the teaching profession (Redecker, 2017). According to the DigCompEdu, there are twenty-two educator-specific digital competencies organized in 6 areas that are focused on different aspects of educators’ professional activities (see table 1).

Table 1. Areas of educator-specific digital competences

<table>
<thead>
<tr>
<th>Sections</th>
<th>Areas of Digital Competences</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Professional Engagement</td>
<td>1.1 Organizational Communication</td>
<td>OC Using digital technologies for communication, collaboration</td>
</tr>
<tr>
<td></td>
<td>1.2 Professional Collaboration</td>
<td>PC and professional development</td>
</tr>
<tr>
<td></td>
<td>1.3 Reflective Praxis</td>
<td>RP</td>
</tr>
<tr>
<td></td>
<td>1.4 Digital CPD (Certified Program Development)</td>
<td>DCPD</td>
</tr>
<tr>
<td>II. Digital Resources</td>
<td>2.1 Selecting Digital Resources</td>
<td>SDR Sourcing, creating and sharing digital resources</td>
</tr>
<tr>
<td></td>
<td>2.2 Creating and Modifying Digital Resources</td>
<td>CMDR</td>
</tr>
<tr>
<td></td>
<td>2.3 Managing, protecting and sharing digital resources</td>
<td>MPSDR</td>
</tr>
<tr>
<td>III. Teaching and Learning</td>
<td>3.1 Teaching</td>
<td>T Managing and orchestrating the use of digital technologies in teaching and learning</td>
</tr>
<tr>
<td></td>
<td>3.2 Guidance</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>3.3 Collaborative Learning</td>
<td>CL</td>
</tr>
<tr>
<td></td>
<td>3.4 Self-regulated learning</td>
<td>SrL</td>
</tr>
<tr>
<td>IV. Assessment</td>
<td>4.1 Assessment strategies</td>
<td>AS Using digital technologies and strategies to enhance assessment</td>
</tr>
<tr>
<td></td>
<td>4.2 Analyzing evidence</td>
<td>AE</td>
</tr>
<tr>
<td></td>
<td>4.3 Feedback and Planning</td>
<td>FP</td>
</tr>
<tr>
<td>V. Empowering Learners</td>
<td>5.1 Accessibility and Inclusion</td>
<td>AI Using digital technologies to enhance inclusion, personalization and learners’ active engagement</td>
</tr>
<tr>
<td></td>
<td>5.2 Differentiation and Personalization</td>
<td>DP</td>
</tr>
</tbody>
</table>
VI. Facilitating Learners’ Digital Competence

<table>
<thead>
<tr>
<th>Competence</th>
<th>Area</th>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1 Information and Media Literacy</td>
<td>IML</td>
<td></td>
<td>Enabling learners to creatively and responsibly use digital technologies for information, communication, content creation, wellbeing and problem-solving</td>
</tr>
<tr>
<td>6.2 Digital Communication and Collaboration</td>
<td>DCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.3 Digital Content Creation</td>
<td>DCCr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.4 Responsible Use</td>
<td>RU</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.5 Digital Problem Solving</td>
<td>DPS</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. Materials and methods

3.1 Participants

The target population of this pre-piloting and item revision study was academic teachers of a few Polish universities (N=103). Ethical approval to conduct the research has been addressed. Participants have been provided with sufficient information about the goals of the study. To protect and respect the personal data provided by participants, the survey tool was anonymous. The data was collected during the 2018-2019 academic year. The sample included 50.5% female respondents, 42.7% male respondents and 5.8% of respondents did not indicate their gender. The age range of the respondents was between 28 and 69 years. Participants’ teaching experience ranged from 1 year to more 40 years.

3.2 Instrument and procedure

In our study, the two-factor TPACK model was used for analysis. As the first step it was preliminary investigated how academic teachers perceive their levels of digital competences regarding different competence areas (table 1). Obtained results were consistent with modifying TPACK theoretical model, and then the possible links between academic teachers’ digital competences in different areas and their demographic (gender, age, previous academic experience and titles and degrees) and professional (using of different digital techniques and resources for teaching) backgrounds were analyzed. In this quantitative survey study, an anonymous questionnaire was designed based on the European DigCompEdu self-assessment tool. All items of the questionnaire were assigned to the appropriate area of digital competences and Cronbach’s alpha item reliability analysis was done. Items with the Cronbach’s alpha score less than 0.75 were eliminated from the study. After checking the properties of the items, an analysis was performed to verify how closely related a set of items are as a group. For that purpose, the reliability of the instrument was calculated. Cronbach’s alpha was computed to measure the internal consistency of the instrument for the Polish sample of teachers.

Finally, university experts evaluated selected items and marked, which of these two factors of the modifying TPACK model particular areas of digital competences represent (e.g. F1PK or F2TK) and how representative items are describing in the model on a 5-point scale. Factors and internal consistency of the TPACK scale are shown in table 2.

3.3 Data analysis

Several quantitative research methods were applied to establish evidence for the validity and reliability of the instrument. Whole instrument with 73 items and each of the six competence areas for internal consistency using Cronbach's alpha reliability technique was assessed. Statistical analysis was carried out as follows. A descriptive analysis between variables of the chosen demographic and professional background and different areas of the digital competences carried out.

4. Results

4.1 Levels of teachers’ digital competences

To reach the first research question we used SPSS 25. Levels of digital competences were established based on percentiles: low level – till 25 percentiles, medium level – between 26 to 75 percentiles, and higher-level from 76 and above (see Table 3).
A pre-pilot diagnostic test of levels of digital competences of academic teachers in the Polish higher educational environment concerning TPACK factors allows to state that about half of the respondents scored a medium level in the examined range for both factors (PK and TK). About a third of respondents (34% F1PK and 30% F2TK) show a low level of digital competences. Polish academic teachers scored lowest on a higher level of digital competences (19% for the F1PK and 16% for the F2TK).

4.2 Relationships between demographics and professional backgrounds and areas of digital competences

To examine the correlations between demographic and professional backgrounds and areas of digital competences the following variables were taken into account: gender, age, years of teaching, titles and degrees (as the demographic background); providing of on-line course; the percentage of teaching time for using digital technologies, using presentation for teaching, watching videos during classes, creating videos for teaching, using online learning environments, using digital quizzes or polls, using interactive apps or games, using digital posters, mindmaps, planning tools, using blogs or wikis (as the professional background).

Taking into account gender, this variable differentiated IML (Mmale = 2.621; SD = .471 vs. Mfemale = 2.386; SD = .588), t(94) = 2.176; p< .05; Cohen’s d = .450) and RU (Mm = 2.298; SD = 0.399 vs. Mf = 2.477; SD = .450), t(94) = -2.065; p< .05; Cohen’s d = .430).

Perception of Information and Technology Literacy (F2TK) among male respondents was higher compared with the perceptions of female participants. Opposite results concern Responsible Use (F2TK) which were higher among female comparing with the male.

Other relationships were analyzed using Person correlations, significant results are summarized in Table 4.

Table 3. Academic teachers’ levels of digital competences
There were significant negative relationships between demographic background variables (age, years of teaching and titles and degrees) and the majority of areas of digital competences. The highest
negative correlations were between age and DCPD (F1PK) and SDR (F2TK), years of teaching and DP (F1PK), and titles and degrees and DCPD (F1PK).

All these results could be explained by the low technological skills of the older generation of academic staff in Poland's higher education environment as well as their reluctance according to digital professional development. At the same time, there were no statistically significant correlations between age and AE (F1PK), FP (F1PK) as well as between years of teaching and CL (F1PK) and FP (F1PK). All of these areas of digital competences were from the first Pedagogical Knowledge TPACK Factor. And obtained results might be clarified by academician teaching experience what allows them to be updated with analyzing evidence to support students, organize collaborative learning as well as use different ways to provide feedback either with the use of digital approaches.

The majority of positive associations were between variables from the professional background and a lot of areas of digital competencies from both TPACK factors (PK and TK). The highest correlations were between SrL and using for teaching online learning environments, digital quizzes or polls, interactive apps or games; AS and online learning environments; FB and online learning environments, and digital quizzes. In the case of TPACK factors, significant positive correlations were between F1PK and such professional background variables as providing on-line courses; the percentage of teaching time for using digital technologies; creating videos for teaching; using online learning environments, digital quizzes or polls, interactive apps or games, digital posters, mindmaps, planning tools, blogs or wikis. All these activities prove academic teachers’ digital pedagogical knowledge and influence positively on the rising of the digital competences’ level. Significant negative correlations were noticed between F2TK and the age of the respondents and positive links with the percentage of teaching time for using digital technologies in class and using blogs or wikis for teaching and learning.

5. Discussion

Investigation of the levels of Polish university teachers’ digital competences based on percentiles shows that near 50% of the respondents are at the medium level in the examined range for both factors (PK and TK), near 30% demonstrated low level and less than 20% are in the higher level of the digital competences. These results might be opposite to the statement about the lower rating of the technological knowledge of the university teachers (Blayone et al., 2018; Castéra et al., 2020; Cubeles et al., 2018).

The second aim was to find links between university teachers’ backgrounds variables and areas of digital competences. On completion of a joint analysis of the responses, the averages of the different areas of digital competences were compared to study whether there were significant differences according to the respondents’ demographic and professional backgrounds.

Perception of Information and Technology Literacy (F2TK) among male respondents was higher compared with the perceptions of female participants. These results proved previous studies conclusions that male teachers are more confident in using computers than their female colleagues (Markauskaitė, 2006); males report higher T-competency than females (J. Koh et al., 2015; Scherer et al., 2017). Opposite results concern Responsible Use (F2TK) which were higher among female comparing with the male. These results might be explained that EFA analysis showed the similar practical significance of the RU for both factors – pedagogical and technological knowledge. So, females in our study demonstrated higher scores in the area which could belong to Pedagogical Knowledge factor. This results confirming the studies which gave similar conclusions (Lin et al., 2013; Luik et al., 2017), and it is in contrast with one of the last study which claims that there is no gender difference in the TPACK perception (Castéra et al., 2020).

Regarding the rest of demographic variables (age, years of teaching and titles and degrees), significant negative correlations have been confirmed (Castéra et al., 2020; J. H. L. Koh et al., 2010; Lee & Tsai, 2010; Luik et al., 2017). This result is also in contrast to previous studies, such as Cubeles (2018) where didn’t confirm differences in the TPACK domains for any age group and Lin et al. (2013) who argued that gender did not have a significant effect on the preservice teachers’ perceptions of pedagogical knowledge and technological knowledge. The highest negative correlations were between age and Digital Certified Program Development (DCPD) area which belongs to the Pedagogical knowledge factor and Selecting Digital Resources (SGR) area from the Technological Knowledge factor. To cope with this negative trend the training modules based on the specific lacks should be organized in each university. Since age is an important factor influencing technology knowledge and to
integrate it into instruction, Learning Activity Types approach (Hofer & Harris, 2010) for building TPACK and differentiating experimented and unexperimented teacher educators would be an interesting framework for developing teacher educators’ TPACK (Castéra et al., 2020).

There were noticed a lot of positive associations between variables from a professional background and different areas of digital competences from both TPACK factors (PK and TK). The biggest positive influence on the higher scores of digital competences was using for teaching online learning environments, digital quizzes or polls, interactive apps or games; providing on-line courses; creating videos for teaching. The values obtained could be due to the intrinsic use of technology in the online format as the professors have to design their course and adapt the contents to an online learning environment and this process itself improves their technological knowledge (Cubeles et al., 2018). Virtual collaboration increases teachers’ opportunities to work with different technologies (Bueno-Alastuey et al., 2018), increases critical thinking, and develops collaboration skills (Kopcha et al., 2014). This coincides with other studies that teacher achievement goal orientation is strongly associated with practices of pedagogical ICT use (Karaseva et al., 2018). Our findings reveal that a video is a key tool in teaching and learning, using video for remote teaching/learning is commonplace in higher education and the group of regular video producers contains a larger proportion of university teachers (Espino et al., 2020).

6. Conclusion

We can conclude that the study survey for investigation of teachers’ digital competencies developed on the base of the TPACK scale connected with the self-assessment tool of the European Competence Framework for the Digital Competence of Educators is valid and useable in the higher education context. The CFA confirmed the two-factor TPACK model of technology and pedagogy areas.

Polish academic teachers seem to have a medium level of digital competences both in Pedagogical as well as Technological Knowledge. Therefore, teacher training and re-training programs for the professional development of university teachers should pay more attention to those elements. Also, it was not surprising that age, years of experience teaching and titles, and degrees of university teachers negatively correlated their Technology Knowledge; however, we did not expect that their evaluations on some domains of Pedagogical Knowledge also be significant.

The present study argues about the positive influence of practical use of different digital technologies for teaching and learning on the rise of all areas of teachers’ digital competences. Specifically aimed at university teachers, the developed and tested survey with satisfactory validity and reliability proposed in this paper may be useful to further explore the TPACK of university teachers in other countries. Moreover, it is the next challenge to discover indicators to qualify the teachers’ professional development in different areas of digital competences.
### Table 4. Correlations for some demographic and professional backgrounds variables

<table>
<thead>
<tr>
<th>Areas of Digital Competences</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 OC</td>
<td>-.202*</td>
<td>.252*</td>
<td>.245*</td>
<td>.269**</td>
<td>.475**</td>
<td>.375**</td>
<td>.360**</td>
<td>.374**</td>
<td>.342**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2 PC</td>
<td></td>
<td>.215*</td>
<td>.199*</td>
<td>.368**</td>
<td>.290**</td>
<td>.337**</td>
<td>.428**</td>
<td>.379**</td>
<td>.323**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.3 RP</td>
<td>-.231*</td>
<td>.242*</td>
<td>.291**</td>
<td>.256**</td>
<td>.317**</td>
<td>.247*</td>
<td>.269**</td>
<td>.230*</td>
<td>.310**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4 DCPD</td>
<td>-.360**</td>
<td>-.246*</td>
<td>-.254**</td>
<td>.241*</td>
<td>.384**</td>
<td>.231*</td>
<td>.297**</td>
<td>.228*</td>
<td>.216*</td>
<td>.297**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.1 SDR</td>
<td>-.304**</td>
<td></td>
<td>.252*</td>
<td>.249*</td>
<td>.264**</td>
<td>.267**</td>
<td>.379**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.2 CMDR</td>
<td></td>
<td></td>
<td>.249*</td>
<td>.264**</td>
<td>.267**</td>
<td>.379**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.3 MPSDR</td>
<td></td>
<td></td>
<td></td>
<td>-.233*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.241*</td>
</tr>
<tr>
<td>3.1 T</td>
<td>.197*</td>
<td>.235*</td>
<td>.239**</td>
<td>.362**</td>
<td>.218*</td>
<td>.235*</td>
<td>.276**</td>
<td>.267**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.2 G</td>
<td>.233*</td>
<td>.297**</td>
<td>.329**</td>
<td>.318**</td>
<td>.314**</td>
<td>.236*</td>
<td>.474**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.3 CL</td>
<td></td>
<td></td>
<td>.228*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.213*</td>
</tr>
<tr>
<td>3.4 SrL</td>
<td>-.196*</td>
<td>-.195*</td>
<td>-.203*</td>
<td>.291**</td>
<td>.316**</td>
<td>.342**</td>
<td>.574**</td>
<td>.598**</td>
<td>.524**</td>
<td>.434**</td>
<td>.309**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.1 AS</td>
<td>-.196*</td>
<td>.331**</td>
<td>.260**</td>
<td>.405**</td>
<td>.573**</td>
<td>.487**</td>
<td>.498**</td>
<td>.360**</td>
<td>.240**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.2 AE</td>
<td></td>
<td>.384**</td>
<td>.409**</td>
<td>.311**</td>
<td>.289**</td>
<td>.390**</td>
<td>.234**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.3 FP</td>
<td>.328**</td>
<td>.248*</td>
<td>.267**</td>
<td>.497**</td>
<td>.510**</td>
<td>.411**</td>
<td>.425**</td>
<td>.428**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.1 AI</td>
<td></td>
<td></td>
<td></td>
<td>.207*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.2 DP</td>
<td>-.248*</td>
<td>-.266**</td>
<td>-.198*</td>
<td>.199*</td>
<td>.268**</td>
<td>.385**</td>
<td>.297**</td>
<td>.247**</td>
<td>.337**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.1 IML</td>
<td>-.118*</td>
<td>.225*</td>
<td></td>
<td>.325**</td>
<td>.388**</td>
<td>.288**</td>
<td>.309**</td>
<td>.319**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.2 DCC</td>
<td></td>
<td></td>
<td></td>
<td>.248*</td>
<td>.213*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.3 DCCr</td>
<td></td>
<td>.197*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.242**</td>
</tr>
<tr>
<td>6.4 RU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.5 DPS</td>
<td></td>
<td></td>
<td>.269**</td>
<td>.216*</td>
<td>.268**</td>
<td>.340**</td>
<td>.278**</td>
<td>.301**</td>
<td>.323**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>---</td>
<td>---</td>
<td>-------</td>
<td>-------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1PK</td>
<td></td>
<td></td>
<td>.326**</td>
<td>.278**</td>
<td>.425**</td>
<td>.576**</td>
<td>.527**</td>
<td>.490**</td>
<td>.504**</td>
<td>.350**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F2TK</td>
<td>.267**</td>
<td></td>
<td></td>
<td>.233*</td>
<td></td>
<td>.248*</td>
<td>.196*</td>
<td></td>
<td></td>
<td>.332**</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 – age; 2 – years of teaching; 3 – titles and degrees; 4 – providing online courses; 5 – the percentage of teaching time for using digital technologies; 6 – using presentation in teaching; 7 – watching videos during classes; 8 – creating videos for teaching; 9 – using online learning environments; 10 – using digital quizzes or polls; 11 – using interactive apps or games; 12 – using digital posters, mindmaps, planning tools; 13 – using blogs or wikis

* correlation is both sides significant on the level 0.05; ** correlation is both sides significant on the level .01
References


Learning Dialogues orchestrated with BookRoll: A Case Study of Undergraduate Physics Class During COVID-19 Lockdown

Vijayanandhini KANNAN\textsuperscript{a*}, Jayakrishnan WARRIEM\textsuperscript{b*}, Rwitajit MAJUMDAR\textsuperscript{c} & Hiroaki OGATA\textsuperscript{c}

\textsuperscript{a}Department of Physics, School of Sciences, GITAM (Deemed to be University), Hyderabad, India.
\textsuperscript{b}NPTEL, Indian Institute of Technology Madras, Chennai, India.
\textsuperscript{c}Academic Center for Computing and Media Studies, Kyoto University, Japan.

\textsuperscript{*}vkannan@gitam.edu, \textsuperscript{*}jkm@nptel.iitm.ac.in

Abstract: With COVID-19 pandemic forcing academic institutions to shift to Emergency Remote Teaching strategies, teachers worldwide are attempting several strategies to engage their learners. Even though existing research in online learning has proven that effectiveness of the online session is more dependent on pedagogical design rather than technology features, most teachers still focus on the intricacies of the technology. In this paper we present the adaptation of an active learning pedagogy - Learning Dialogue (LeD), for an undergraduate physics classroom. We used the eBook reader BookRoll to orchestrate an LeD along with the support of a video conferencing tool and a dashboard that provides immediate input on the engagement. The adaptation of the strategy utilized the appropriate affordance of each tool available in generating an engaging session for the students. Comparison of the student artefacts (memos in BookRoll) between regular face-to-face classroom session and online session indicated that there is a sustained engagement in the online class. Preliminary qualitative analysis also shows that the students were posing good conceptual clarifications/questions aligned with the session agenda.

Keywords: Emergency Remote Teaching, Covid19 Lockdown, Learning Dialogue, BookRoll eBook Reader, LAView Analysis Dashboard, Physics Education Research

1. Introduction

The sudden transition to Emergency Remote Teaching (ERT) (Hodges et.al., 2020) has left teachers and students with very little time to prepare for the changes that are ahead of them. Existing research in online and distance learning has already pointed out the need for promoting self-reflection, self-regulation and self-monitoring as an important factor in determining positive learning outcomes among students (Means et. al., 2009). This means that the key idea during this transition is for faculty to focus more on the pedagogical skills rather than technology skills for ensuring success of the online course experience (Shieh, Gummer, & Niess, 2008; Garrison, Cleveland-Innes, & Fung, 2010). However, with known barriers of attitude and skill deficit among teachers for online learning practices (Keengwe & Kid, 2010), it is really difficult to expect a large number of teachers to smoothly transition into an effective online facilitation mode during these times.

Due to COVID-19 pandemic, shifting from the face-to-face (f2f) to an online teaching has now become mandatory. With several concerns reported, it still remains challenging to constructively engage students in the f2f teachings (Felder & Brent, 2009). In an online teaching, it is even difficult to bring-in the classroom culture and maximize learning (Millikan, 1996). There are several challenges reported pertaining to online teaching. Few of them are social integration, low student engagement behavior, high drop-out rates, etc. (Levy, 2007). The problems are compounded for novice teachers, as they could lack recommendations for the best online pedagogical practices, adequate time to adapt to the new tool and institutional support towards providing the infrastructural facility. Scouring the existing research to make sense and then adopt those learning seems to be difficult in the current situation. Major hindrances experienced from the perspectives of online learners may be, the apprehension towards adapting new technology tools due to various levels of learner competency, feel of isolation and being out of comfort f2f learning zone (Felder & Brent, 2009). Despite the best intentions to provide an effective learning experience, most of the teachers feel apprehended to quickly switch over to the online learning. In order
to attain a better online student engagement, it is desirable for teachers to implement an active-learning (AL) online pedagogy along with the available technological affordances that would appropriately fit to their teaching context (Amy, Janet, Evelyn, & Sarah, 2012). An effective implementation of online teaching and learning practices need much detailed investigation that addresses the issues as perceived by both the teachers and learners.

We propose the use of active learning strategies that are closer to the existing face-to-face practices of the teacher along with the use of technology that supports reflection-in-practice as a possible solution to help faculty in smoothly transitioning into online educational practices. We provide an example of orchestrating pedagogy of Learning Dialogues (LeD) with technology affordances offered by the tool BookRoll. The case being described in the current paper is that of an undergraduate physics class for engineering students in India. The instructor has utilized the MOODLE and BookRoll features available as part of the Technology-enhanced Evidence-based Education and Learning (TEEL) project for more than a year. In the earlier report (Vijayanandhini & Sai Preeti, 2019), we show that the adoption of AL strategies in flipped learning method using the TEEL tools had been effective to better engage the students and enhance learning. In the current paper, we examine the pedagogic strategy called the LeD orchestrated with BookRoll in two different teaching contexts: (i) regular f2f class during the pre-lockdown (abbreviated as f2-LeD) and (ii) purely online conducted during lockdown with the help of synchronous meeting tool - GoTo Meeting™ (abbreviated as on-LeD). Both the f2-LeD and on-LeD orchestrations included the flipped learning provided within the TEEL platform as the learning management system (LMS). The study investigates the following research questions:

RQ1: What is the difference in engagement of students while participating in f2-LeD and the on-LeD methods?
RQ2: What is the variation in the performance of students in assessments conducted after the f2-LeD and on-LeD, when following the LeD orchestrated with BookRoll pedagogy for similar topics?

2. Literature Review

Physics educational researchers have shown that the simple conceptual acquisition from a limited context may not be sufficient to solve real world problems (Van Heuvelen, 1991). To diversify and deepen the conceptual acquisition, implementation of AL pedagogy is often emphasized. With several known benefits of AL strategies, many higher educational institutions have shifted away from the traditional lecturing. On the other hand, there is an increasing demand to develop technology tools and online platforms to cater the needs of creating different range of interactive teaching activities. In this section, first we highlight different AL strategies reported in the literature for the f2f class and online teaching. Next, we describe the AL pedagogical model adapted in our present study to create the flipped online contents. Finally, we introduce an LMS based learning analytics platform that has been designed with many optimal dashboard features which assisted the seamless orchestration of our active learning pedagogy.

2.1 Active Learning Pedagogies for f2f and online teaching

Several AL strategies are reported for f2f teaching to promote physics education at a tertiary level degree programs (Mintzes, Walter, 2020). Some of cooperative learning techniques includes peer instruction (Crouch, Mazur, 2001), think-pair-share (McTighe, & Lyman, 1988), jigsaw method (Aronson, & Patnoe, 1997), etc. In all of these strategies, the students are allowed to discuss with the peers to learn concepts while the teacher facilitates the process. In online teaching, there are technology affordances that helps to leverage the effectiveness of AL. However, adapting AL along with the technology adaptions to navigate and engage students within the online setting are still being explored (Nurul, Martin, & Frances, 2015). The researchers have shown how to appropriately shape the instructional design of online lecture videos to effectively engage learners using the click-stream data tool (Lin, Aiken, Seaton, 2017). Few reports highlight that the conceptual videos should be short and avoid abrupt transitions (Kim, Guo, Seaton, 2014). Whipp & Lorentz (2009) suggests to maintain an effective student interaction by asking a challenging question and then, provide a timely and concise feedback to those seeking help.
2.2 Pedagogical foundations of current work: LCM Model

The Learner-centric MOOCs (LCM), ‘a prescriptive model consisting of a set of guidelines, activity formats and actions for MOOC creators’ has been proposed (Murthy et.al., 2018). The model emphasizes interactive activities rather than traditional information transfer. Thus, it maintains a learner-centric pedagogy as its main orchestration. As shown in the Figure 1, the pedagogical basis of the LCM model consists of four active-learning structural components as follows:

(i) Learning dialogue (LeD), is the first element of LCM that promotes concept acquisition through learner interaction. The key design feature of an LeD is the reflection spot, a place for a learner to express prior conceptions, perform micro-practice or reflect. LeD videos (less than 15 minutes) can be created using interactive (H5P) video tool having multiple choice questions embedded within it to engage students.

(ii) Learning by Doing (LbD), an online quiz with customized and constructive feedback. LbD activities are formative assessment activities that provide learners with frequent opportunities to practice and apply their learning. It helps learner towards the goals of concept attainment, immediate application or integration of knowledge.

(iii) Learning Extension Trajectories (LxT), advanced resource materials to diversify student’s learning.

(iv) Learner Experience Interaction (LxI), a discussion forum activity to cultivate a structured discussion forum interaction through focus questions. The focus questions drive and keep the discussion centred on a specific topic.

2.3 TEEL Infrastructure: BookRoll and LAViEW dashboard

We orchestrated our course on the Technology-enhanced and Evidence-based Education and Learning (TEEL) platform (Ogata et al., 2018) and incorporated LAViEW (Majumdar et al., 2019). Figure 2 shows the four major components of TEEL. The learning behavior sensor captures the learner’s and teacher's interaction data during the session. It offers a LMS, e.g. MOODLE, that integrate other e-learning tools. For instance, we used BookRoll, an e-book reader and LAViEW, the associated learning analytics dashboard (Ogata et al., 2015, Majumdar et al., 2019). BookRoll allow students to read digital contents such as lecture slides or materials that are shared by instructor. It has a feature like red or yellow markers to highlight some parts of the text that are important or difficult to understand. Additionally, students can add memos to remember important points, annotate doubts or comments. They can bookmark pages to access them easily while reviewing the content. These actions are recorded and then can be viewed by the instructor to understand the reading habits of students in the learning analytics dashboard LAViEW (Majumdar et al., 2019). Literature reports suggest that the reading behavior of students can be used to visualize class preparation...
and review patterns (Fu, Shimada, Ogata, Taniguchi, et al. 2017). LAViEW contains various panels of visualized indicators for monitoring and plays a central role to assist and identify problems in the teaching-learning scenario based on analysis of the visualized indicators. Both teachers and students can access these learning tools. Thus the TEEL infrastructure integrates the features of the eReader, LMS and Dashboard within a single service so that teachers can seamlessly move across the technology.

3. Description of online orchestration strategy

In the present study, we utilized the TEEL platform to flip the learning content during both the f2-LeD and on-LeD strategies. Contents of every module were chunked into multiple sub-topics. The flipped contents for each sub-topic was created following the structure of LCM model: LeDs as short videos followed by LbD quiz, extended resources as LxT and the LxI forum activity. The f2f pedagogy allows a teacher to intersperse his/her lecture with regular activity to engage the learner with the concept. During online teachings, the instructor has to carefully choose the format of the content and activities based on the available technology resources. To adapt the pedagogy of LeD orchestrated with BookRoll into a technology mediated setting in the current context, the teacher designed the orchestration as shown in Figure 3 below.

(i) **Session with GoTo Meeting and LeD video**

Deliver the content through lecture mode in the GoToMeeting (a web conferencing tool) online sessions using the LeDs videos and BookRoll pdf materials of the MOODLE as shown by steps 1 & 2 in the pedagogical flow diagram (Figure 3). Design an explicit spots at pause points called “clarification spot” (as shown in Figure 4 a) in the BookRoll, where the students are required to reflect on the session till that time and post a query or ask for clarification in the form of a memo. Encourage students to note down their doubts/clarifications through the memo function of BookRoll tool during the online lecture as shown in Item number 4 in the Figure 3. The memos so created are seen by the instructor through the LAView dashboard of TEEL platform (item 5 in the pedagogical flow) and is followed by verbal explanation/clarification given by the teacher instantly during the same sessions (as shown as item 6 in the pedagogical flow).
(ii) **Reflection spot activity with BookRoll**

To elicit misconceptions and ensure that students have achieved conceptual understanding, these clarifications are followed by the next set of reflection spots, 1 to 3 (given as practice questions) that required students to conceptually reflect or do micro-practice on the content learnt till this time (Item 7) [screenshot image shown in Figure 4 b]. The answers that a student comes up with after this reflection/micro-practice are gathered through memos or as answers to BookRoll or LbD quizzes (Item 8 in the pedagogical flow).

![Book Roll Memo Bank to submit Doubts](image)

*Figure 4. Screenshot of (a) Clarification Spot and (b) Reflection Spot activity created in BookRoll tool during the on-LeD method.*

(iii) **Feedback sharing**

Share the LAView dashboard screen during the online session to summarize the concepts (Item 9 in pedagogical flow) or augment the understanding of solutions to the reflection spot problems.

4. **Research Methods**

4.1. **Research Design**

The instructional design, namely, the LeD orchestrated with BookRoll is adapted within two different teaching contexts: (i) regular f2f class during pre-lockdown (f2-LeD) and (ii) fully online during the COVID-19 lockdown (on-LeD). Both the f2-LeD and on-LeD included the flipped teaching in TEEL system – the LCM based activities primarily addressing the out-of-class component, while the BookRoll based LeD activity addressing the in-class component. The pedagogical flow as discussed in the Figure - 3 only differed in the timeline of implementation for both the f2-LeD and on-LeD methods. In case of on-LeD, all the steps from 1 to 9 were integrated within the same online sessions as an immediate post-teaching activity. Whereas, in case of the f2-LeD, the students were required to note their doubts/clarification (step - 4) in BookRoll tool provided in the LMS during the out-of-class learning, while summarizing (step - 9) were followed in the next f2f teaching session (as shown in the Table-1).

<table>
<thead>
<tr>
<th>Pedagogical flow in steps</th>
<th>f2-LeD</th>
<th>On-LeD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step-1. Content delivery</td>
<td>Lecture during regular f2f</td>
<td>Lecture in GoToMeeting</td>
</tr>
<tr>
<td>Step-3. Clarification Spots</td>
<td>flip phase (out-of-class)</td>
<td>online immediate after Step-1</td>
</tr>
<tr>
<td>Step-7. Reflection Spots</td>
<td>f2f session arranged in the computer lab once in a week</td>
<td>online immediate after Step-3</td>
</tr>
<tr>
<td>Step-9. Summarizing</td>
<td>Next f2f class sessions</td>
<td>online immediate after Step-7</td>
</tr>
</tbody>
</table>
4.2 Course and Participants

The study was implemented in freshman undergraduate engineering (B.Tech) students with the specialization (major) in electronics and communication at GITAM (Deemed to be University), India. A total of 58 students were offered the Engineering Physics (19EPH131) course during the semester-II. The students were provided with an individual log-in to TEEL platform to access the learning activities of the flip phase. Same instructor taught the course during both the f2-LeD and on-LeD teaching. 19EPH131 consisted of five modules, out of which first three and a half modules were taught during the pre-lockdown and the last one and a half module was taught during the lockdown. To address the RQs, we had selected few target topics from the modules taught during the f2-LeD and on-LeD teaching methods. The topic equivalence was checked with the similarities in pre-requisite knowledge required to learn.

4.3 Data collection and Analysis overview

We attempt to answer the RQ-1 with three different data sets as collected from the LAView analysis dashboard of TEEL platform: (i) count of specific student logs (only ‘viewed’) for LeD activities and total LbD quiz attempts from the TEEL platform during the course period, (ii) total count of student memo responses collected during the ‘clarification spots’, answers to the reflection spot problems for the target topics, (iii) analysis of quality of memos annotated in BookRoll tool for the target topics. Students activity report recorded by the TEEL platform such as log data can be extracted by the MOODLE administrator. These data offer insights on the learning progress of students in the flipped learning contents. Therefore, we used them as a metric for student engagement [Rosalina & Rodolfo, 2017]. As we had three materials for LeDs of the target topics, we averaged the total student logs per single material. Then, we compare the log data for f2f and fully online teaching. Secondly, the quality of memos was analyzed keeping the checkpoints as: (a) ‘Whether the memos are relevant to the key concepts dealt’, (b) ‘Whether memos have been created after carefully reflecting on the session contents’. The RQ-2 was addressed by comparing the mean scores of online quizzes conducted after the BookRoll activities of f2-LeD and on-LeD methodologies. It is important to note that we compare the post-test quiz scores of two different cohorts of participants implemented with different target topics. To justify the comparison, we select the post-test scores of only those questions that measured similar level of thinking skills (say ‘Understand’ or ‘Apply’). Paired T-test analysis was then carried out on these scores to compare the effectiveness of learning. We used iSAT, a visual analytics tool to understand the transition pattern of student learning (Majumdar & Iyer, 2016).

5. Results

5.1 Student Engagement

Creating memos was integral part of both the f2-LeD and on-LeD sessions. Table 2 shows the number of student memos during the f2-LeD and on-LeD sessions for two different target topics. We observed that the total count of memos collected as the clarification spots in BookRoll tool were higher during on-LeD as compared to f2-LeD. It was observed that during synchronous on-LeD teaching sessions, the students proactively engaged in the BookRoll activities to create memos. Further, the instructor could instantly clarify those doubts raised before initiating the reflection spot activity. In contrast, during the f2-LeD, regular engagement of student in the flipped contents, to read and annotate memos was not observed. In case of f2-LeD, there was a lag between the topics covered during the f2f session and those flipped topics for which the student submitted their memos in BookRoll. This led to difficulty for the instructor to clarify the doubts during the f2f sessions on concepts for which the students had misconceptions. Analysis of the correct answer counts for the reflection spot questions in BookRoll indicated that the students could answer most of questions correctly. In fact, number of students who could correctly answer the questions were much higher for those different topics dealt-in on-LeD as compared to the f2-LeD (Table 2). The number of student responses for consecutive reflection spot questions from 1 to 3 decreased in both the teaching phases.
The preliminary qualitative analysis of the memos posted during the online setting indicates that the learners were able to focus on the key concepts of the online lecture. The memos collected during online sessions showed that clarifications raised by students were more relevant to the session agenda. There were also memos pointing to concepts that were going to be covered in the next sessions. For example, the memo: “If the electric force would balance the magnetic force, then in this case Lorentz force would be zero” shows that the student had a misconception in understanding the balances between basic forces acting on charged particles in the Hall effect experiment. This clarification spot was addressed instantly during the session, which also helped students to better understand the subsequent topics of the same session. Further, in one another memo, students raised memos like “Please explain in detail of why the photo diode works only in large depletion condition and how such wide breakdown voltage is created in large depletion zones” showed that the students had paid attention to the topics being taught online and had put sufficient thinking on contents before raising the doubt.

5.2 Student performance

To analyse the performance of the students, we first compare the transition of the scores obtained by learners in the post-tests as shown in the Figure 5. We see that there is an effective upward transition for learners. The transition pattern is created by considering three strata of scores (high - , medium - , low- ) that are obtained by students in the post-tests (Majumdar & Iyer, 2016). This transition pattern shows that 36 learners scored high score in the post-test conducted after the on-LeD session, compared to the 25 who obtained high score after f2-LeD. Out of this 36 high scorers, 15 students had initially scored a medium/low score in the post-test after the f2-LeD. Now to further verify statistically whether this is a promising trend, we compare the mean score of the group of students in both the post-tests. Post-activity test analysis of selected questions showed a higher mean score (M = 9.48 out of 10) for on-LeD teaching as compared to the f2-LeD (M = 8.37 out of 10). As shown in the Table 3, the paired T-test analysis of post-test scores indicated a statistically significant p-value (p < 0.05). This shows that the on-LeD method led to a better student academic performance as compared to f2-LeD method.

6. Discussion

Due to the sudden lockdown condition caused by the Covid-19 pandemic, many institutes and Universities had to switch over to the digital mode of teaching and learning within short notice. Therefore, it may be difficult to achieve the desired learning outcomes in the end-semester examinations, as switching to online teaching is not easy and demands unique skill sets from the teachers. To adapt to this sudden transition, we had implemented a teaching strategy called LeD orchestrated with BookRoll to conduct the synchronous online teaching using the GoToMeeting web conferencing tool. Analysis of MOODLE log data showed that we could effectively engage students during the online teaching. The students engagement in the BookRoll activities showed substantial improvement as compared to f2f class. Qualitative analysis of the memos submitted during online teaching inferred that the students paid more attention to the contents and asked doubts reflecting on the sessions. Presently, the f2-LeD and on-LeD strategies are implemented in two different teaching
contexts. Therefore, we emphasize that there is a limitation to compare both the strategies taking evidences only from the quantitative data analysis. The pedagogical design of the learner-centric MOOCs has proven to be effective for conducting online faculty development programs and MOOC courses for the diverse and a massive scale of learners (Warriem, Murthy, Iyer, 2016; Veenita, Gargi, Murthy, et., 2018]. The authors have earlier reported on contextual improvisation and data-driven validation of blended LCM model adapted for a small and less diversified group of learners in regular engineering physics course (Vijayanandhini & Sai Preeti, 2019; Kuromiya, Majumdar, Warriem, & Ogata, 2019]. In the current study, we adapted the pedagogical design of LeD, which is the first learning component of the LCM model. The pause points of the LeD design was provided as ‘clarification spots’, which helped the students to reflect or collate their doubts using the Bookroll tool. The feedback or doubts as collected by the LAView dashboard features of TEEL platform helped the Instructor to instantly address the portions where the students needed further clarifications. It also assisted to add new content to augment better understanding of the topics. Importantly, the clarification spots helped to break the monologue in the content delivery in online. Following the clarification spots, students were prompted with the BookRoll based reflection spot activity that required to solve and answer at least two to three reflection spots (as practice questions) using the memo functions. This allowed students to do micro-practice immediately of those concepts being taught within the same online sessions. The instructor could then review all the solutions of the problems and reasonings using the LAView dashboard, to further help students to engage better even during the post teaching activity. The reflection spots were then followed by the summarisation of diverse view-points and a final closure of the pedagogical learning design of the LeD. Thus, the teacher could create a better engaging and interactive online sessions utilizing the technological features (BookRoll, LAView dashboard) available in the TEEL platform.

The overall student attendance behavior during on-LeD sessions was only about 50 to 60%. Thus, the students were instructed to learn asynchronously from the flipped LCM materials such as the
LeD videos, BookRoll based activities and the LxT resources. We found that more students submitted the solved problems provided as the asynchronous assignments through the BookRoll tool. This, in turn, facilitated those learners who could not attend the synchronous GoToMeeting sessions due to unavoidable reasons like non-availability of gadgets or network connectivity issues. The analysis of post-test after the BookRoll activity for both the f2f and fully online sessions indicated a statistically significant difference in the mean scores. Thus, we show that the academic performance of students did not hamper or dip due to the sudden transition to online teaching. We adopted a similar pedagogical flow during the f2-LeD where the instructor had used the f2f class (instead of web conferencing) for delivering content and clarifying the initial doubts. However, the response of the students to get engaged and learn from the flipped activities was not completely positive. It was observed that they spent more time on exploring the technology tool rather than focusing on the content. Additionally, since there are restrictions on bringing own devices inside the classroom, all the BookRoll activities were conducted in a computer lab by carefully adjusting the time slots with other instructors. It can be noticed that there is an advantage of adjusting session, especially, in context to the teacher-mediated f2f classrooms settings. However, with most of the learning being transferred to online mode after the lockdown, the logistical limitations (of lab availability) were now removed and the instructor could focus more on the pedagogical design of the online activity. Presently, we could utilize the new dashboard tools (LAView & BookRoll) of TEEL platforms, to collect the real-time feedback and conduct an interactive online learning session, which is otherwise not possible with any other MOODLE platforms, particularly during the COVID-19 pandemic lockdown.

7. Conclusions

In the case study as discussed here, we see that transition to online settings have provided the instructor more flexibility to incorporate careful pedagogic strategies in the teaching-learning process. The available technology infrastructure also supports the instructor in the process by providing real-time data of student engagement so that further time could be invested in addressing student queries and clarifications. The teacher has carefully selected a strategy that is closer to a regular f2f strategy that both teacher and students are more exposed to. As Shieh et.al (2008) and Garrison et.al (2010) recommends, this has allowed the instructor to carefully design pedagogical actions to be done using the various technology features (GoTo Meeting, MOODLE, BookRoll and LAView) both by the teacher as well as the student during the online settings. The familiarity of the tool to both the teacher and student (through prior use in regular f2f setting) has helped to minimize the time required to adjust to the usage of tool in an online setting. This is an important factor to consider while selecting tools for online teaching-learning usage.

Acknowledgements

This research was supported by NEDO Special Innovation Program on AI and Big Data 18102059-0, JSPS KAKENHI 16H06304, 20K20131 and SPIRITS 2020 of Kyoto University.

References


When is flipped classroom more effective and why it flops

Khe Foon HEW

The University of Hong Kong, Hong Kong SAR
kfhew@hku.hk

Abstract: A specific form of blended learning model – the flipped classroom approach – is increasingly being adopted by many education institutes around the globe. Yet, many researchers and practitioners remain divided over whether flipped classroom is really an improvement over the traditional classroom. This paper is the first to synthesize all available meta-analytic information on the effectiveness of the flipped classroom approach on student cognitive outcomes, and provide the currently best available evidence on the optimum use of flipped classroom. Cognitive outcomes refer to the domain-specific knowledge (e.g., facts, concepts) of a subject. This paper also summarizes the key reasons why flipped classroom fails. Using a meta-synthesis approach to examine 19 flipped classroom meta-analytic studies on student cognitive outcomes, involving 1,126 empirical studies and more than 85,000 flipped and 90,000 non-flipped participants, this paper found positive significant effects, ranging from small to medium effect size with a median of 0.42, favouring the use of flipped classroom in achieving better student cognitive outcomes. Current evidence suggests flipped classroom is more effective when formative assessment (e.g., quizzes) are used. Flipped classroom appears to be equally effective among learners from different educational levels. However, findings concerning the possible moderating effect of flipped classroom implementation duration and across subject disciplines were inconclusive. This paper also identifies several key student- and instructor-related challenges factors that could lead to flipped classroom failure through a comprehensive meta-synthesis of 16 systematic reviews and meta-analyses. Practical suggestions for alleviating these challenges are discussed. Overall, the findings presented in this paper can provide useful recommendations to help practitioners design more optimal flipped classroom lessons.

Keywords: Flipped classroom, flipped learning, meta-synthesis, effect, learning outcomes

1. Introduction

Educators have long been intrigued with the potential of technology to help enhance student learning. One such technology supported instructional approach that has captured the attention of many K-12 and university educators around the globe is the flipped classroom model. In a flipped classroom model, students learn lecture materials before class usually through some form of technology-supported means such as video recorded lectures, and online exercises. Students then spend the in-class time completing some form of active learning tasks such as group discussions, and independent work to apply what they have learned under the supervisor of the instructor. Advocates of the flipped classroom approach believe that increasing active learning opportunities can help students learn the lecture materials better. Others such as Abeysekera and Dawson (2015) postulate that the self-paced nature of pre-class activities in a flipped classroom can reduce cognitive load which can in turn theoretically enhance learning outcomes.

1.1 Problem statement

Nevertheless, not every educator is sold on the idea of flipped classroom. A survey of 290 European universities revealed that only 15% of institutions found the flipped classroom approach to be “fully useful” (Gaebel & Zhang, 2018). Critics claim that it is difficult to get students on board the flipped classroom model, and that the model only benefits students from families with more resources.
Although extensively studied over the years, there is still debate about the effectiveness of flipped classroom in improving learner outcomes (Strelan, Osborn, & Palmer, 2020; Zainuddin, Haruna, Li, Zhang, & Chu, 2019). On one hand, some primary studies show that flipped classroom can induce significantly better student learning performance (e.g., Schultz, Duffield, Rasmussen, & Wageman, 2014; Tsai, Shen, & Lu, 2015). On the other hand, other primary studies found flipped classroom fails to improve student learning performance (e.g., traditional learning is just as effective as flipped learning) (e.g., DeSantis, Van Curen, Putsch, & Metzger, 2015; Yong, Levy, & Lape, 2015). Moreover, a recent randomized controlled trial experiment published in a Massachusetts Institute of Technology discussion paper that involved a total of 1,328 students across 80 economics and mathematics classes found that flipped classroom exerted no long-term effects on student learning, but exacerbated the achievement gaps between white and black or Hispanic students when compared to the traditional lecture group (Setren, Greenberg, Moore, & Yankovich, 2019).

1.2 Prior meta-analyses of flipped classroom

In order to address the problem of conflicting results, researchers in recent years have begun to conduct meta-analyses of flipped classroom primary studies. Meta-analyses can help estimate the size and consistency of effect sizes of flipped classroom implementations across different studies and quantify the variance when effect sizes vary. A meta-analysis can also help identify possible factors that modify that effect (Gurevitch, Koricheva, Nakagawa, & Stewart, 2018).

To the best of our knowledge, the earliest meta-analysis on flipped classroom was published in 2016. Since then the number of flipped classroom meta-analyses have skyrocketed. As many as 19 meta-analyses examining the effects of flipped classroom versus non-flipped classroom on student cognitive outcomes have been conducted so far (see Results). Student cognitive outcomes, usually accessed through tests and exams, refer to the domain-specific knowledge of a subject such as concepts, theories, and facts (Klein, Kuh, Chun, Hamilton, & Shavelson, 2005). Rather than doing yet another meta-analysis of flipped classroom, this study calls for a temporary moratorium on new meta-analyses. Instead, this study aims to synthesize the quantitative results of all currently available meta-analyses in order to provide a general picture of what we currently know about the effects of flipped classroom. Put another way, this study aims to capture the main essence of what the extant body of literature says about the quantitative impact of flipped classroom on student cognitive outcomes and what factors (if any) may render flipped classroom more effective.

1.3 Contribution of the present study

In this study, I performed a meta-synthesis of all 19 flipped classroom meta-analyses on student cognitive outcomes. The research approach of meta-synthesis springs from the interpretive paradigm of naturalistic inquiry (Guba, 1978; Noblit & Hare, 1988) and is aimed at “understanding and describing key points and themes contained within a research literature on a given topic” (p. 4). The steps of meta-synthesis typically include the following (Walsh & Downe, 2005): (a) search for articles, (b) make decision on inclusion of articles, (c) appraise studies (judge the quality of the included articles, for example identify the potential confounds), (d) analyze studies (for example determine how studies are similar or different through a compare and contrast exercise), and finally (e) synthesize findings. Meta-synthesis is an interpretive, rather than an aggregating method that aims to integrate the findings from various studies (Walsh & Downe, 2005). Although meta-syntheses are traditionally used to examine quantitative meta-analytic studies, several researchers (e.g., Strobel & van Barneveld, 2009; Wilder, 2014) have expanded the use of meta-syntheses to synthesize qualitative meta-analytic studies.

This study also summarizes the main student-related challenges that can undermine the use of flipped classroom. Understanding the key student-related challenges can help educators take the necessary steps to address them. A meta-synthesis of 16 systematic reviews and meta-analyses revealed two key student-related reasons why flipped classroom may flop. Please note that some meta-analyses reported student-related challenges. Practical suggestions, gleaned from relevant empirical studies, to mitigate these challenges are discussed. Overall, the findings presented in this
paper can provide useful recommendations to help practitioners design more optimal flipped classroom lessons. The following research questions guided the present study:

**Research question 1:** What are the general findings of previous meta-analyses concerning the overall effect of flipped classroom on student learning outcomes?

**Research question 2:** What factors contribute to more effective use of flipped classroom as reported by previous meta-analyses?

**Research question 3:** What are the key student- and instructor-related challenges that can undermine the use of flipped classroom as reported by previous systematic reviews, as well as meta-analyses?

2. Method

2.1 Search strategy

I searched for as many systematic reviews, as well as meta-analytic articles as possible. To do this, I performed a search of major educational databases such as Academic Search Complete, British Education Index, ERIC, MEDLINE, Teacher Reference Center using the following search terms ("review" OR "synthesis" OR "meta-analysis") AND ("flip*" OR "invert*") AND ("class*" OR "learn*"). The date of publication remained open for the initial search. Besides the search of academic databases, I also conducted a Web search (e.g., Google Scholar, Google) and screened the reference lists of relevant articles. All articles must be written in English.

2.2 Data extraction

To address RQ1: “What are the general findings of previous flipped classroom meta-analyses concerning the overall effect of flipped classroom versus non-flipped classroom on student learning outcomes”, I extracted the following data from each eligible meta-analysis article: (a) the type of subject discipline, (b) participant grade level, (c) the number of primary empirical studies, and (d) effect size data concerning student learning outcomes. Student learning outcomes must be measured by tests or exams. Meta-analysis articles that relied solely on subjective measures such as student perceived learning outcomes or non-cognitive outcomes (e.g., student perceived satisfaction) were excluded.

To address RQ2: “What factors contribute to more effective use of flipped classroom as reported by previous meta-analyses?”, I extracted the moderator analyses results from each eligible meta-analysis article.

To address RQ3: “What are the main student- and instructor-related challenges that can undermine the use of flipped classroom as reported by previous flipped classroom systematic reviews, as well as meta-analyses (if relevant)?”, I adopted the grounded approach. The first step was an initial reading of all qualitative data (e.g., student and instructor comments about the challenges of flipped classroom) to obtain an overall idea of the data, and to generate relevant emerging codes. Similar codes were organized into themes. To enhance the consistency of coding, several exemplary quotes that clearly illustrate each constructed theme were identified.

3. Results

3.1 Articles reviewed

Figure 1 shows the PRISMA flowchart that illustrates the entire article screening process. The initial academic databases search resulted in 2,912 records. Eighteen additional records were identified by searching the Web. After removing duplicates, 2,527 remained. The titles and abstracts of the remaining 2,527 records were screened. Many records were excluded because they were irrelevant to the purpose of the present study (e.g., FLIP as a therapeutic target in cancer). Subsequently, 49 full-text records were assessed for eligibility. Of these 49 full-text records, 14 were excluded because they did not focus on student cognitive outcomes, or the specific challenges of flipped classroom.
implementation. Ultimately, 35 records consisting of 19 meta-analyses and 16 systematic reviews were included in the present meta-synthesis.

3.2 **RQ1:** “What are the general characteristics and findings of previous meta-analyses concerning the overall effect of flipped classroom versus non-flipped classroom on student cognitive outcomes”

The set of 19 included meta-analyses with a list of education context, number of primary studies, number of participants, effect size metric, and overall average effect size is presented in Table 1.

<table>
<thead>
<tr>
<th>Meta-analysis</th>
<th>Education context</th>
<th>Primary Studies (N)</th>
<th>Participant (N)</th>
<th>Effect size metric</th>
<th>Overall mean effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algarni (2018)</td>
<td>All levels</td>
<td>34</td>
<td>All: 8,598 (separate samples not reported)</td>
<td>SMD</td>
<td>0.27, p &lt; .05</td>
</tr>
<tr>
<td>Chen et al. (2018)</td>
<td>Higher education</td>
<td>46</td>
<td>FC: 4,526, NFC: 4,828</td>
<td>SMD</td>
<td>0.35, p &lt; .05</td>
</tr>
<tr>
<td>Cheng et al. (2019)</td>
<td>All levels</td>
<td>55</td>
<td>FC: 3,727, NFC: 4,185</td>
<td>g</td>
<td>0.19, p &lt; .05</td>
</tr>
<tr>
<td>Gillette et al. (2018)</td>
<td>Higher education</td>
<td>5</td>
<td>FC: 685, NFC: 710</td>
<td>g</td>
<td>0.37, p &lt; .05</td>
</tr>
<tr>
<td>Hew &amp; Lo (2018)</td>
<td>Higher education</td>
<td>28</td>
<td>FC: 2,295, NFC: 2,420</td>
<td>SMD</td>
<td>0.33, p &lt; .05</td>
</tr>
<tr>
<td>Hu et al. (2018)</td>
<td>Higher education</td>
<td>11</td>
<td>FC: 742, NFC: 742</td>
<td>SMD</td>
<td>1.06, p &lt; .05</td>
</tr>
<tr>
<td>Kang &amp; Shin (2016)</td>
<td>All levels</td>
<td>36</td>
<td>Not reported</td>
<td>SMD</td>
<td>0.53, p &lt; .05</td>
</tr>
<tr>
<td>Karagöl &amp; Esen (2019)</td>
<td>All levels</td>
<td>55</td>
<td>FC: 2,210, NFC: 2,400</td>
<td>g</td>
<td>0.57, p &lt; .05</td>
</tr>
<tr>
<td>Låg &amp; Sæle (2019)</td>
<td>All levels</td>
<td>271</td>
<td>FC: 23,856, NFC: 27,372</td>
<td>g</td>
<td>0.35, p &lt; .05</td>
</tr>
<tr>
<td>Lo &amp; Hew (2019)</td>
<td>All levels</td>
<td>29</td>
<td>FC: 2,590, NFC: 2,739</td>
<td>g</td>
<td>0.29, p &lt; .05</td>
</tr>
<tr>
<td>Lo et al. (2017)</td>
<td>All levels</td>
<td>21</td>
<td>FC: 1,620, NFC: 1,564</td>
<td>g</td>
<td>0.30, p &lt; .05</td>
</tr>
<tr>
<td>Orhan (2019)</td>
<td>Higher education</td>
<td>13</td>
<td>FC: 328, NFC: 323</td>
<td>g</td>
<td>0.74, p &lt; .05</td>
</tr>
</tbody>
</table>
Shi et al. (2019)  Higher education  33  FC: 3,674  NFC: 3,273  SMD 0.53, *p < .05*  
Strelan et al. (2020)  All levels  198  FC: 15,641  NFC: 18,037  *g* 0.50, *p < .05*  
Tan et al. (2017)  Higher education  29  FC: 1,896  NFC: 1,798  SMD 1.13, *p < .05*  
van Alten et al. (2019)  All levels  114  FC: 12,017  NFC: 12,661  *g* 0.36, *p < .05*  
Zhang (2018)  All levels  28  FC: 7,847  NFC: 5,624  *g* 0.42, *p < .05*  
Zheng et al. (2020)  All levels  95  All: 15,386 (separate samples not reported)  *g* 0.66, *p < .05*  
Zhu et al. (2019)  K-12  25  FC: 1,739  NFC: 1,579  SMD 0.56, *p < .05*  

Note: FC refers to flipped classroom; NFC refers to non-flipped classroom; SMD refers to standardized mean difference, *g* refers to Hedges’ *g*.

All 19 meta-analyses included two-group comparison design studies such as quasi-experiments, historical control cohort, and experiments. In a quasi-experiment or historical control cohort, the flipped class would typically be treated as the experimental group, while the non-flipped class as the control group. Only two meta-analyses (Hu et al., 2018; Tan et al., 2017) restricted inclusion to randomized control trials (RCTs), which is a common recommendation in meta-analysis (Reeves et al., 2008), albeit difficult to operationalize in education due to the paucity of RCT studies. Out of the 18 meta-analyses, 16 employed the random effects model in their computation of the effect sizes since the conditions that could affect learner outcome, including frequency of lessons flipped, student population, and course level, may differ among studies in the analysis (Gurevitch & Hedges, 1999). One meta-analysis (Orhan, 2019) employed the fixed effects model, while another (Algarni, 2018) did not explicitly report which model was used.

All 19 meta-analyses examined the effect sizes of primary studies obtained from peer-reviewed journals. Several meta-analyses included the analysis of additional primary studies from unpublished sources (e.g., Chen et al., 2018; Cheng et al., 2019; Karagöl & Esen, 2018; Liu et al., 2018; van Alten et al., 2019). Seven of the 19 meta-analyses conducted some form of methodological quality assessment on the primary studies they reviewed (Chen et al., 2017; Gillette et al., 2018; Hew & Lo, 2018; Hu et al., 2018; Liu et al., 2018; Tan et al., 2017; Xu et al., 2019). It is interesting to note that these meta-analyses all focus exclusively on disciplines within health professions education (e.g., nursing, medicine, pharmacy). One possible reason for the prevalence of quality checking on health professions education research as compared to other fields is the global movement for quality in medical education. Five meta-analyses employed the Cochrane risk-of-bias tool to assess each primary study to determine whether the study yielded a low risk of bias (unlikely to seriously alter the results), a high risk of bias (seriously weakens confidence in the results), or an unclear risk of bias (Gillette et al., 2018; Hu et al., 2018; Liu et al., 2018; Tan et al., 2017; Xu et al., 2019). The other two meta-analyses used the Medical Education Research Study Quality Instrument (MERSQI) (Hew & Lo, 2018), and the Effective Public Health Practice Project Quality Assessment (EPHPP) Tool (Chen et al., 2018).

Overall, all 19 meta-analyses concluded that flipped classroom improved learners’ cognitive outcomes significantly better than non-flipped classroom. The basic findings of these meta-analyses varied from weak to strong support of flipped learning. The overall average effect sizes for cognitive outcomes ranged from 0.19 to 1.13, with a median of 0.42. The two largest effect sizes (1.06 and 1.13) came from studies conducted exclusively with Chinese nursing students (Hu et al., 2018; Tan et al., 2017) reported in Chinese-medium publications.

3.3 RQ2: “What factors contribute to more effective use of flipped classroom as reported by previous meta-analyses?”

Next, we turn to the question, when is flipped classroom more effective? To do this, I examined the moderators analyzed in the previous meta-analyses, and their results (e.g., *Q* data, and *p* values). A
majority of meta-analyses commonly examined the use of quizzes in flipped classroom, the educational level of participants, the subject disciplines, the study duration or implementation duration, and the types of publication. There are, of course, other moderating factors that were analysed but we will focus only on these aforementioned factors due to page constraint of this article.

We observe empirical support showing that flipped learning appears to be more effective when quizzes are used before and/or during class time. Evidence for this statement is provided in five meta-analyses (Hew & Lo, 2018; Lag & Saele, 2019; Lo & Hew, 2019; Lo et al., 2017; van Alten et al., 2019). A moderator analysis of health professions education studies indicated that the effect size was significantly higher when quizzes were employed at the start of face-to-face class lessons (p = 0.02, Hew & Lo, 2018). A meta-analysis of flipped mathematics education studies also revealed that the effect size was significantly higher when instructors used a structured formative assessment such as a quiz at the start of face-to-face lessons compared to instructors who did not (p = 0.013, Lo et al., 2017). Van Alten et al. (2019) similarly found that the use of quizzes in flipped learning showed a significant difference (p = 0.044) compared to studies where quizzes were not included.

The use of flipped classroom appears to be equally effective among learners from different educational levels. No significant effect size difference among elementary, high school, undergraduate and graduate learners were found (Cheng et al., 2019; Karagöl & Esen, 2019; Lag & Saele, 2019; Lo & Hew, 2019; Orhan, 2019; Tan et al., 2017; van Alten et al., 2019). The effect on cognitive outcomes did not appear to be moderated by different publication types (Cheng et al., 2019; Kang & Shin, 2018; Orhan, 2019; van Alten et al., 2019).

Results concerning the possible moderating effect of flipped classroom implementation duration were mixed. For example, on one hand, van Alten et al. (2019) found no significant effect size difference among different durations (e.g., 1-10 weeks, > 10 weeks). Cheng et al. (2019) similarly found no significant effect size among durations of one semester, and one semester or more. On the other hand, Zheng et al. (2020) found that the effect of flipped learning in social sciences was significantly higher than in implementation duration of 5-8 weeks as compared to 2-4 weeks, 9-24 weeks, and more than 24 weeks. One possible reason for this is that different meta-analyses used different categories to classify the implementation duration – hence making it difficult to make useful comparisons. Results concerning the impact of flipped learning on different subject disciplines were also mixed (e.g., Cheng et al., 2019; Shi et al., 2019; van Alten et al., 2019). One possible reason for this is that different meta-analyses used different categories to classify the subject disciplines – hence making it difficult to make meaningful comparisons across different meta-analyses.

3.4 RQ3: “What are the key student- and instructor-related challenges that can undermine the use of flipped classroom as reported by previous flipped classroom systematic reviews, as well as meta-analyses?”

A meta-synthesis of 16 systematic reviews and meta-analyses found several key student-related and one key instructor-related challenges that could lead to flipped classroom failure.

**Student-related key challenge 1: Unwilling to prepare for class.** Not all students are willing to complete the pre-class work due to two key reasons or factors. First, students are unhappy with the perceived extra workload in terms of time and effort imposed by the pre-class activities (Akçayır & Akçayır, 2018; Al-Sammarraie et al., 2019; Betihavas et al., 2016; Bond, 2020; Brewer & Movahedazarhouligh, 2018; Kraut et al., 2019; Lo & Hew, 2017; Lo et al., 2017; Njie-Carr et al., 2017; O’Flaherty et al., 2015; Ramnanan & Pound, 2017; Turan & Akdag-Cimen, 2019; Ward et al., 2018; Zainuddin & Halili, 2016; Zhang, 2018). Second, may students still prefer the traditional classroom format because they do not like or perceive the value of active learning inherent in flipped classroom (Akçayır & Akçayır, 2018; Betihavas et al., 2016; Bond, 2020; Brewer & Movahedazarhouligh, 2018; Karabulut-Ilgu et al., 2018; Lo & Hew, 2017; Lo et al., 2017; Vanka et al., 2020; Ward et al., 2018; Zainuddin et al., 2019; Zhang, 2018).

**Student-related key challenge 2: Not engaged with pre-class work.** First, students are not able to get immediate help or feedback while they study at home (Al-Sammarraie et al., 2019; Bond, 2020; Akçayır & Akçayır, 2018; Lo & Hew, 2017; Lo et al., 2017; Ramnanan & Pound, 2017; Zainuddin et al., 2019). Second, the video lectures are not interesting to watch (Akçayır & Akçayır, 2018; Al-Sammarraie et al., 2019; Bond, 2020; Karabulut-Ilgu et al., 2018; Lo & Hew, 2017; Lo et al., 2017; Njie-Carr et al., 2017; Zainuddin & Halili, 2016). Students are unmotivated to watch the videos due to
three main reasons: the videos are too long, the audio quality is poor, and the perception that videos are not as important as worksheets.

**Instructor-related key challenge: Unwilling to try out flipped classroom.** Instructors’ reluctance of implementing flipped classroom appear to stem from three main factors: unfamiliarity with the technological tools (e.g., video recording technology), unhappiness with the perceived extra workload before and during class, and familiarity with the traditional lecture format (Al-Samarraie et al., 2019; Akçayır & Akçayır, 2018; Betihavas et al., 2016; Karabulut-Ilgu et al., 2018; Kraut et al., 2019; Lo & Hew, 2017; Lo et al., 2017; Njie-Carr et al., 2017; O’Flaherty et al., 2015; Ward et al., 2018; Zainuddin & Halili, 2016). By far, the most commonly reported factor was instructor unhappiness with the perceived extra workload. The actual time needed for an instructor to prepare flipped course materials can be nearly six times more than traditional course preparation (Wanner & Palmer, 2015). In addition, during in-class sessions, an instructor may need to serve many students requesting assistance at the same time (Karabulut-Ilgu et al., 2018).

4. **Discussion**

Probably the main practical implication we can draw here is that flipped classroom is worth implementing. Overall, meta-analytic evidence suggests that flipped classroom is more effective than traditional classroom in improving learner cognitive outcomes. All meta-analyses reported positive significant effects, ranging from small to medium effect size with a median of 0.42, favouring the use of flipped classroom. Compared to non-flipped classroom, flipped classroom provides students with more than one exposure to the course materials. Students are first exposed to the course materials during the pre-class activity. Students engaged with the course materials again later during the in-class session. Multiple exposure to course materials can help improve student understanding of the lesson.

Future implementations of flipped classroom should incorporate formative assessment such as short reviews or quizzes since the use of reviews tends to significantly increase the effect size of student cognitive outcomes. These reviews may consist of specific instructor’s generated questions to assess student learning based on the pre-class materials. The use of a review enables an instructor to determine students’ possible factual or conceptual misunderstandings about the content materials. If students’ misunderstandings are identified, the instructor can provide the necessary remedial action such as reviewing the pre-class materials or changing the in-class teaching plans to specifically correct the misconceptions.

Despite the overall advantage of flipped classroom over the traditional approach, it is important to note that not all flipped classroom implementations are smooth. This paper identifies several student- and instructor-related challenges that could diminish the benefits of flipped classroom. I shall discuss some practical recommendations gleaned from relevant empirical studies to alleviate these challenges in the following sections.

**How can practitioners mitigate student unwillingness to complete the pre-class work?** To recall, students are unwilling to do the flip pre-class work because they perceive the pre-class work as extra workload, and that they prefer the passivity of traditional teacher-lecture format. In order to deal with the former, it is important for the flipped classroom instructor to retain the overall workload hours as in its traditional format. Lo and Hew (2017) suggest that instructors first estimate the total time typically required for the students to complete the homework that is traditionally done outside the classroom. Instructors can then use this time estimation as a reference when designing their out-of-class learning activities of flipped classrooms.

A more difficult challenge is to address the students’ preference for the relatively passive learning of a traditional classroom. Flipped classrooms require the students to spend their in-class lessons on active learning activities (e.g., group discussion) instead of merely listening to a teacher’s lectures. Students tend to view active learning activities negatively because they dislike spending the required additional cognitive effort to complete the activities (Deslauriers, McCarthy, Miller, Callaghan, & Kestin, 2019). Students also perceive they learn less than in passive classes (Deslauriers et al., 2019). However, students’ perceptions were faulty because students in active learning classrooms in reality learned more (Deslauriers et al., 2019). Since the success of flipped classroom hinges on student ‘buy-in’ of active learning, it is important that efforts be made to explain to students the advantages of active learning in the classroom. Early intervention by the instructors such as explaining clearly the value of increased cognitive effort and convince students that they can benefit
from active learning (Deslauriers et al., 2019) to be helpful in fostering positive student attitude toward active learning.

**How can practitioners address the problem of student disengagement with the pre-class work?** Unlike a traditional classroom, students in a flipped classroom environment cannot ask their teacher while watching the instructional videos. Previously some studies (e.g., Bhagat et al., 2016) suggest the use of online discussion forum for students to post their questions and discuss with other people. However, the asynchronicity of a forum or email introduces a time lag between postings and replies. This time lag could discourage students from posting their comments (Hew, Tang, Lo, Zhu, 2018). To overcome this problem, teachers can use a mobile instant messaging app such as WhatsApp or WeChat for students to seek help. MIM apps such as WhatsApp and WeChat allow users to engage in quasi synchronous communications on their mobile phones. In times of urgent communication needs, many students may only have their phones available. When a MIM message arrives, a notification will automatically show up on the user’s phone screen, which encourages timely response (Hew et al., 2018). Findings from a study (Hew et al., 2018) suggest that MIM app can be viable platform to encourage students to communicate with other people when they need to seek help or feedback from other people.

Students can also be disengaged when they watch video lectures particularly when the videos are too long. To deal with the problem of long videos, instructor should consider segmenting it into shorter clips of about 6 min long (Guo et al., 2014). Besides the issue of video length, practitioners should also be concerned about the quality of the audio in the video, as well as how the instructor presents the course materials on videos. Some students perceive that videos are not as important as worksheets. For example, a few students complained about watching videos in Snyder et al. (2014) that “I feel like I’m just reading and listening to facts”. It is hard for students to retain any information that they learn from a video simply by watching it once, especially when the video contains new information (Lam, Hew, Jia, 2020). Having a worksheet to fill in can help focus learner attention to the video’s details, and possibly watching it multiple times (Lam et al., 2020).

**How can practitioners mitigate instructors’ unwillingness to try out flipped classroom?** The support from IT staff is essential in helping instructors to implement flipped classroom (Critz & Knight, 2013). Not every instructor is an experienced user of flipped classroom. They may thus be inexperienced in handling the technical problems of video production. Gaughan (2014) recalled the experience of creating his first instructional video. He stated that it was a painless experience with the assistance of IT staff. In fact, “With low-cost computer-based video capture capabilities becoming more readily available, capturing, editing and posting digital video recordings is a realistic option for anyone” (Albert & Beatty, 2014, p. 419). For example, applications such as “Microsoft Office Mix” for Windows users and “Explain Everything” for iOS users are some user-friendly tools of producing instructional videos. However, it would be better if universities can provide training of using these new technologies.

Many instructors also lamented that it was time consuming to prepare flipped classroom materials especially the instructional videos. However, it is not necessary to flip the entire course at one go (Naccarato & Karakok, 2015). In fact, instructors can “start, and proceed, at a reasonable pace” (Snyder, Paska, & Besozzi, 2014, p. 314) by working on two to three topics every year (Lo & Hew, 2017). Starting small with these topics can also enable teachers to gain experiences of implementing flipped classroom. Instructors can also utilize existing video clips from YouTube and Khan Academy. Nevertheless, it is important to note that although instructors may shorten the video preparation time by editing existing video resources, many of these existing videos do not offer personalization or specificity unlike custom video made by the instructor (Alpert, 2016). Instructor self-created videos are found to be more appealing than videos that do not feature the instructor (Bond, 2020). Therefore, it would still be better for the instructors to develop their own videos in the long run. Even though a significant amount of start-up effort is required to create flipped classroom resources, these resources can be reused in subsequent semesters, which makes the preparation of a flipped course cost-effective in the long run.

### 5. Conclusion

This study is the first to synthesize the available meta-analytic information on the effectiveness of the flipped classroom approach on student cognitive outcomes. It also summarizes the key reasons why
flipped classroom flops. Here I conclude by presenting several implications for practice and future research. From a practical perspective, this paper identifies a useful set of meta-analyses for educators, policy-makers and researchers to draw upon when thinking about the flipped classroom approach. This paper also provides useful recommendations to alleviate the key challenges in order to help educators design more optimal use of flipped classroom. In future work, it would be interesting to examine and compare some of the design strategies for the in-class activities implemented in the flipped classroom studies with largest effect sizes versus lowest effect sizes. Future research should also investigate how the conventional flipped classroom approach can be transformed into fully online flipped classroom to support student learning during emergency school closure periods. Presently, a majority of the primary studies examined were only one semester long in duration. Short-term studies carry the risk of the novelty effect. Learners may become bored with the same flipped classroom approach over time which might diminish their desire to use it. Longitudinal studies are sorely needed to examine whether and how learners’ engagement with flipped classroom changes over time.

Acknowledgements
The research was supported by a grant from the Research Grants Council of Hong Kong (Project reference no: 17610919).

References
Due to page constraints, references for the meta-synthesis articles in Table 1 are not listed here. Please request them from the author.
Gaughan, J. E. (2014). The flipped classroom in world history. History Teacher, 47(2), 221-244

620


Can Computer-Based Learning Environments Mitigate Large Class Size?

Ryan S. BAKER\textsuperscript{a,b}, Aishah AL YAMMAHI\textsuperscript{c}, Joe EL SEBAALY\textsuperscript{c}, Ali NADAF\textsuperscript{c}, Tina KAPP*\textsuperscript{c}, & Seth ADJIEI\textsuperscript{d}

\textsuperscript{a}Teachers College Columbia University, United States of America
\textsuperscript{b}University of Pennsylvania, United States of America
\textsuperscript{c}Alef Education Consultancy LLC, United Arab Emirates
\textsuperscript{d}Northern Kentucky University, United States of America

*ryanshaunbaker@gmail.com

Abstract: In this paper, we study the impact of class size in several schools in the United Arab Emirates using a computer-based learning environment as part of their regular classroom practice. Within this study, we compare larger and smaller classes in terms of both learning and time on task. Classes in the same school are compared directly, and a matching quasi-experimental study is also conducted, comparing students across schools. We find no evidence that the larger classes perform more poorly than smaller classes.

Keywords: Class Size, Computer-Based Learning Environments, Implementation

1. Introduction

In many countries, policy-makers and the general public tend to believe that smaller classes are better for children. In Western Europe and North America, many studies seem to suggest that smaller class sizes are associated with better learning outcomes (Glass, 1982; Shin & Chung, 2009). Others have argued that different teaching practices reduce the importance of class size in East Asia (Stigler et al., 1982; Blatchford et al., 2016). Although there has been less research on this topic in the rest of the world, there is some evidence that smaller class size is associated with better learning outcomes in the Middle East, North Africa, and sub-Saharan Africa (Altinok & Kingdon, 2012). However, the desire to decrease class size clashes with worldwide challenges in recruiting and retaining highly-skilled teachers (Madalinska, 2018). This problem has led many to call for improving teaching practices and increasing teacher ability, rather than reducing class sizes (Stigler et al., 1982; Blatchford et al., 2016).

There is now evidence that computer-based learning environments (CBELEs) can promote enhanced pedagogies that may scale better than traditional approaches to teaching (Miller et al., 2015). The majority of students using these systems work at their own pace, receiving help and feedback from the learning system; teachers focus their time and attention on the specific students that are struggling at any given moment. This strategy has the potential to enable master teachers to take on a larger number of students. However, to the best of our knowledge, there has not yet been research on whether the class size remains associated with student outcomes when students learn from CBELEs in class.

We study this question in the context of an initiative that began in the United Arab Emirates (UAE) in 2017. A public girls’ school piloted using computer-based learning to mitigate the impact of teacher absenteeism, bringing two grade 6 classes together into one larger room as needed. As a result of this experience, the school designed a larger pilot in the fall term of 2018. Two Grade 7 classes for English, Science, and Math were combined into a single class of 60 students. At the same time, the computer-based learning initiative expanded to 6th–8th graders at 57 other local schools in the UAE.

In this paper, we conduct quasi-experimental comparisons to analyze the impact of doubling class size on student time on task and achievement. We compare the students in the large class both to another class in the same school, and separately match the students to comparable students in two other, larger schools. The hypothesis of this study is that larger classrooms using a well-implemented computer-based learning curriculum will not have poorer student learning than smaller classrooms.
2. Methods

2.1 Learning System

We study this difference in class size in the context of the Alef computer-based learning environment. By the 2018-2019 school year, Alef was in use by over 25,000 6th-8th grade students in public sector schools in the UAE, spanning 6 core subjects: Mathematics (English), Science (English), English Language Arts (English), and Arabic Language Arts (Arabic), Social Studies (Arabic), and Islamic Studies (Arabic). Alef follows the UAE Ministry of Education’s national standards, aligned to elements of CCSS (Common Core State Standards) and NGSS (Next Generation Science Standards).

Alef lessons cover specific sets of skills through an instructional process of Explore, Apply, Relate. Students acquire their skills by moving from tasks associated with the initial recall and recognition of foundational concepts and procedures to their application and extension. Within a lesson, the platform detects students' weaknesses through continuous formative assessment. It analyzes student performance on summative assessments as well, and identifies skills that the students are lacking. The assessment-performance data drives automated remediation and practice, and teachers receive real-time data to target support to students with learning differences and needs for offline intervention.

Alef’s content includes both content instruction (in the form of multimedia, videos, and text) and skills application, as shown in Figure 1. Skills practice takes the form of sets of problems to complete. Teachers are also supported in this process through the provision of a series of offline experiential learning kits, usually consisting of manipulatives, simulations, and other hands-on activities. All data collected in this study was drawn from the English language subjects.

![Figure 1. Alef video instruction (left) and assessment of problem-solving skill (right).](image)

2.2 Participants

Data was obtained for three public seventh-grade classes in girls’ schools in Abu Dhabi, UAE. The schools studied in this paper were selected for similar demographics, socioeconomic status, and academic performance profiles, and the students are majority Emirati.

In school A, students were divided into two classes: a large class of 61 students (observation data was obtained for all students; some test data was unavailable for 6 students), and a smaller class of 31 students (test data was obtained for all students; 4 students were absent during the observation days). Students were assigned to different classes by each school; random selection was not used. In schools B and C, 8 additional smaller classes participated, with a total of 204 students (ten absent during the observation days). Each student in the study used Alef as their primary curriculum throughout the course of the school year. As with school A, student-level demographic data is unavailable. Student-level demographic data was unavailable, creating limitations for applying algorithms such as propensity score matching to compare conditions.

2.3 Procedure

We obtained the following data for each of the students in our sample:

Two mathematics diagnostic tests, given in October 2018 and May 2019, using the benchmarked math computer-adaptive test STAR-MATH. On average, students performed 0.62 academic years better on the second test than the first test. STAR-MATH provides two different types of test
scores: Criterion-referenced scores (performance relative to absolute criteria) and Norm-referenced scores (performance relative to other students). In this study, we use Grade Equivalent (GE) Criterion-referenced scores for both diagnostic tests, indicating a grade level for a student in terms of the average knowledge a student would be expected to have in the subject at that grade. The GE score has a range from 0 to 13. These Grade Equivalents are in terms of a US context, but are appropriate for use in the UAE due to the commonalities between the topics taught in each country. Students in the UAE and US need not reach exactly the same point at the same time as students for this measure to differentiate each student’s overall progress.

- Final mathematics examination grade comes from a national exam to measure pupils’ knowledge in all public schools, conducted by the UAE Ministry of Education (MOE).
- Classroom observation data for the students in the sample. BROMP (Baker Rodrigo Ocumpaugh Monitoring Protocol; Baker et al., 2020) was used to collect observations of whether students were on-task or off-task, and whether they were experiencing frustration, boredom, engaged concentration, delight, or confusion. BROMP is a momentary time-sampling method where trained observers make holistic assessments of learner behavior and affect based on a combination of facial expression, posture, and other factors. BROMP has been used in research and practice by over 160 researchers in 6 countries.
- Interaction data from within the Alef platform.

For performance, we focused on mathematics, as diagnostic tests were only available for this subject. For time on task and affect, our observations (three sessions per student) cut across Mathematics, English, and Science, due to the applicability of the method in all cases.

3. Results

3.1 Performance on Exams

Overall, students appear to improve over the course of the year. Across all groups, students improve from an average of 4.45 (SD=1.69) at the first diagnostic test to an average of 5.17 (SD = 2.26) at the second diagnostic test, t(286)=7.94, p<0.001. This gain in performance has an effect size (Cohen’s D) of 0.47, in the range that Hattie (2012) refers to as the “zone of desired effects” and – in Hattie’s view – beyond what even an expert teacher can produce without special practices or programs.

Taking the groups individually, students in the large class had a statistically significant gain from the first diagnostic exam to the second diagnostic exam, t(54)= 3.92, two-tailed p<0.001, 0.93 average gain, Cohen’s D = 0.53. In the smaller class in the same school, the difference in performance between the two exams was not statistically significant, t(27)=0.927, two-tailed p=0.36, 0.14 average gain, Cohen’s D = 0.18. The average gain was significantly higher for students in the large class than students in the smaller class in the same school, t(80.47)=2.79, p=0.007, assuming unequal variances.

Students in the smaller classes in different schools had a statistically significant gain from the first diagnostic exam to the second diagnostic exam, t(203)= 6.83, two-tailed p<0.001, 0.74 average gain, Cohen’s D = 0.48. The average gain was not statistically significantly higher for students in the large class than students in the smaller classes in different schools, t(77.67)=0.75, p=0.46.

As one possible confound, it appears that students were somewhat stronger at the time of the first diagnostic test in the large class (M=5.41) than in the smaller classes in the same school (M=4.11, t(80.71)=4.18, p<0.001) or the smaller classes in the other schools (M=4.24, t(78.90)=4.29, p<0.001). We consider how to address this possible confound in the next section.

On the final examination, students in the large class (M= 76.1, SD=16.6) were outperforming the smaller class in the same school (M=63.0, SD=21.0), t(55.87)=2.28, p=0.026, D=0.69, and students in the smaller classes in the other schools (M=67.4, SD=16.2), t(105.42)=4.87, p<0.001, D=0.53.

Overall, then, our findings do not suggest that being in a larger class hurt student performance.

3.1.1 Matching Analysis

Given that students in the large class had better pre-tests than students in the two comparison conditions, there is some risk that our results are due to a confound. We can address this possible confound, at least in part, through a matching analysis, where we match students in the large class to
similar students in the smaller classes at other schools and then compare the matched students. In doing so, it would be ideal to have full demographic information enabling us to use a sophisticated statistical method such as inverse probability of treatment weighting or propensity score matching. Unfortunately, this information was not available, so we match students based on their performance on the first diagnostic test. We conduct this matching process 1000 times to see how robust the results are to the details of the match. Within this matching process, every student in the large class was represented in the matching and a subset of students in the smaller classes in other schools were matched to the students in the large class, without replacement. Specifically, we randomly shuffle the order of students in the large class, and then for each student in the large class, we find the student closest to them within the smaller classes, removing the students from the data set after matching (i.e. a student cannot be matched twice). In cases where multiple students in the smaller classes had pre-test scores equidistant to the student in the large class being matched, a student was randomly selected to create a matched set.

After matching based on the first diagnostic tests, the data sets are almost identical in terms of that variable, with an average of 5.411 in the larger class and an average of 5.431, across runs (cross-run SD of average = 0.067), in the matched cohorts. The standard deviations for the first diagnostic test are also almost identical, 1.820 for the larger class and an average standard deviation of 1.856 across runs for the matched cohorts (cross-run SD of SD = 0.136).

On the second diagnostic test, the larger class averaged 6.344 and across runs the matched cohort averaged 6.432 (cross-run SD of average = 0.081). This led to very similar levels of gain between the two groups: 0.933 for the larger class and 1.001 for the matched cohort. The difference in gain between the two groups was not statistically significant in any run.

However, the larger class had substantially better performance on the final exam, 70.82, than the average final exam performance of students in the matched cohort, 62.65 (cross-run SD of average = 0.154). The difference in final exam performance between the two groups was statistically significant in 997 of 1000 cases. The average t value across runs was 2.81 (p = 0.007); the standard deviation of the t value across runs was 0.0148. a statistically significant difference, paired t(54)=3.59, p<0.001, average Cohen’s D across runs =0.466 (cross-run SD of average = 0.028).

Hence, even after we control for pre-test differences through matching, we still find that the students in the large class obtain higher final examination scores than the students in the smaller classes.

### 3.2 Affect and Time on Task

The first step to conducting BROMP observations in a new country is to create and refine a locally-appropriate coding scheme (Baker et al., 2020). The finalized coding scheme is tested through an inter-rater reliability round where two coders label student behavior and affect separately but

### Table 2

**Performance by type of class, for matched cohorts. Averages across runs given, with cross-run standard deviations of averages in parentheses. Cases where large school’s values are statistically significantly higher are in boldface.**

<table>
<thead>
<tr>
<th>Type of Class</th>
<th>First Diagnostic Test</th>
<th>Second Diagnostic Test</th>
<th>Gain</th>
<th>Final Examination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large</td>
<td>5.41 (0.07)</td>
<td>6.43 (0.08)</td>
<td>1.00 (0.15)</td>
<td>62.7 (0.15)</td>
</tr>
<tr>
<td>Smaller (matched cohort)</td>
<td>5.41 (0.07)</td>
<td>6.43 (0.08)</td>
<td>1.00 (0.15)</td>
<td>62.7 (0.15)</td>
</tr>
</tbody>
</table>
simultaneously. Three coders conducted observations at this study’s schools. The coders labeled 1,747 observations of up to 20 seconds duration, in 10 classes in 3 schools, including the large classroom, the smaller classroom in the same school, and 8 more comparison classrooms in the two other schools. The coders established inter-rater reliability at School A, two months prior to this study’s data collection. Coders 1 and 2 obtained Kappa of 0.91 for both behavior and affect. Coders 1 and 3 obtained Kappa of 0.72 for behavior and 0.71 for affect.

We compare the proportion of affect and behavior between the larger classroom and the smaller classrooms – focusing on engaged concentration, frustration, boredom, and off-task behavior -- using a two-sample t-test (assuming unequal variance), and using a Benjamini & Hochberg (1995) post-hoc control to adjust for conducting multiple comparisons. The students in the large class displayed significantly more frequent engaged concentration (76%) than the students in the other schools (51%), t(118.96)=6.23, p<0.001, adj α = 0.00625, and also displayed marginally more frequent engaged concentration (76%) than the students in the smaller class in the same school (60%), t(39.56)=2.14, p=0.038, adj α = 0.025. The students in the large class displayed significantly less frequent boredom (5%) than the students in the other schools (15%), t(165.80)=-4.38, p<0.001, adj α = 0.0125. However, there was not a significant difference in boredom between the large class (5%) and the smaller class (2.3%), within the same school, t(72.82)=1.15, p=0.252, adj α = 0.0375. The students in the large class displayed significantly less frequent frustration (0.4%) than the students in the other schools (2.4%), t(243.87)=-2.58, p=0.010, adj α = 0.01875. However, there was not a significant difference in frustration between the large class (0.4%) and the smaller class (1.2%), within the same school, t(33.62)=-0.65, p=0.519, adj α = 0.0435. No significant or marginally significant differences were seen between classes in terms of off-task behavior.

Table 3

Proportion of affect and behavior by type of class. Boldface indicates statistically significantly worse than large class; Italics indicate marginally worse than large class.

<table>
<thead>
<tr>
<th>Type of Class</th>
<th>Engaged Conc.</th>
<th>Boredom</th>
<th>Frustration</th>
<th>Off-Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large</td>
<td>76.0%</td>
<td>5.1%</td>
<td>0.4%</td>
<td>11.0%</td>
</tr>
<tr>
<td>Smaller (same school)</td>
<td>60.3%</td>
<td>2.3%</td>
<td>1.2%</td>
<td>12.6%</td>
</tr>
<tr>
<td>Smaller (other schools)</td>
<td>51.4%</td>
<td>15.4%</td>
<td>2.4%</td>
<td>6.5%</td>
</tr>
</tbody>
</table>

3.3 Usage Data

We can also compare students between the larger and smaller classes in terms of their usage of the Alef system, shown in Table 4. There were not statistically significant differences in terms of the time-on-task between classes, though the difference between the larger class and smaller classes in other schools approached significance, t(94)=1.60, two-tailed p=0.113 assuming unequal variance. There were also no statistically significant differences between groups in terms of the average length of each learning session. Furthermore, we found that there were no statistically significant differences in the number of days the system was used at home by students in each group (based on the day of the week and time of day).

4. Discussion and Conclusions

The results of this study give no evidence for negative impact on the students in the 60-student
classroom. The 60-student classroom exhibited similar or better classroom time on task and academic performance as students in 30-student classrooms. Overall, students in the 60-student classroom appeared to perform acceptably relative to students in other classrooms.

These results do not conclusively demonstrate that there are no negative impacts to having larger classrooms when using a computer-based learning environment. There are several limitations to the existing study. For one thing, given the relatively small sample, it is possible that unaddressed confounds may have artificially enhanced the performance of the students in the larger class. For example, it is possible that the students in the large class were more prepared than students in the other classes, even beyond the differences addressed by controlling for differences in the first diagnostic test.

Another possibility is that the teacher in the large class was more skilled than the teachers in the other classes, producing the positive results seen, or that differences in teachers’ pedagogies and implementation strategies may have compensated for having larger classes. Fine-grained details of implementation and pedagogy can be key to the effectiveness of computer-based learning environments in classrooms (Feng et al., 2014) – as such, we recommend that future work studying the impacts of class size in schools using computer-based learning environments should collect observational data on teachers’ pedagogical strategies and implementation decisions, to help understand the learning and behavior seen in classes of different sizes.

Beyond these possible confounds, the sample is too restricted to draw general conclusions. Our current sample only involves girls-only schools; it is not clear that this pattern will hold in boys-only or mixed-gender schools. It is also unclear whether these results will be seen in other countries, or for other computer-based learning environments. Despite these limitations, though, this paper’s findings suggest that the negative impacts of large class size may be mitigated through the use of computer-based learning environments. Given this finding’s potential economic impacts, it warrants further study.

Acknowledgements

We thank Guzzelle Shahid, Rand Muhsen, and Xin Miao for their assistance in study implementation.

References


Experiences with e-learning as a challenge for the effective training of future generations of teachers

Łukasz TOMCZYK*
Institute of Educational Studies, Pedagogical University of Cracow, Poland
*tomczyk_lukasz@prokonto.pl

Abstract: The paper presents experiences with e-learning among students of educational studies in the biggest Polish pedagogical university that prepares future teaching staff. The research was conducted in May and June 2019 in a group of 450 students (average age=22.660, with st. dev.=4.232). The goal of the study was to present the set of indicators that coexist with the use of distance education platforms by Polish pre-service teachers. Through triangulation of the research tools, eight survey questionnaires of high internal coherence (0.925) were used. Based on the data collected, it was noticed that the respondents have little experience in applying different forms of distance education. Last year, only 8.4% of the students were using paid platforms to develop their interests and knowledge. It was also observed that students who are highly active online (who use different e-services very often or often) show much more interest in online training. Financial status and gender do not determine student engagement in distance education. Participation in one form of e-learning very often coexists with other forms. The key factor indicator of involvement in paid and free e-learning courses is respondents’ activity in cyberspace.

Keywords: teachers, pre-service teachers, e-learning, digital literacy, needs, experiences, Poland

1. Introduction

The last two decades have brought a series of new insights into e-learning. For years, the number of publications on the conditions of distance education at different levels has been growing systematically. These changes have been enforced by the development of the information society in which many offline activities are transposed into cyberspace. An exemplification of this are the increasingly advanced solutions being created to support e-learning or blended learning (Oyelere et al., 2018). In many cases, modern education is also imposed by modernisation processes within institutions, which are introduced intentionally (by following complex recommendations and the principles of the development of e-learning in certain countries) as well as randomly. This last category becomes particularly important when transferring certain activities from the offline into the online world becomes a necessity. An example of such circumstances is the global transformation of stationary teaching into e-learning due to the COVID-19 pandemic (Ting et al., 2020; Pyżalski, 2020). Intentional activities introduced over many years (research, implementation of new platforms, deployment of existing mediated learning environments, digital education methodology improvements) have resulted in e-learning becoming an integral part of academic education and some forms of professional development or formal (school-based) learning. Despite these circumstances, there are many questions regarding effective e-learning, be it planned or forced on institutions by unexpected conditions (Heba et al., 2014).

 Appropriately-prepared teaching staff play a crucial role in the popularisation of e-learning. It is teachers who, in the first place, are expected to operate all of the effective and available means, forms and methods that support e-learning and teaching, and to do so with ability and confidence. In the information society, some of the main tools are, undoubtedly, online applications and websites. However, in order to use the most recent solutions, one needs to have certain resources which allow an understanding of how new media can be used effectively, as well as skills to introduce these solutions into one’s didactic and educational activities. Education is being modernised through computerisation (in this case, the introduction of e-learning) and this is taking place as a result of several important
factors (Tomczyk et al., 2019). Leaving aside the global circumstances, some of these important indicators are environmental and individual conditions. The first category is less dependent on the teacher and is in the first place connected with the institution he or she works in. Support for teachers, regular introduction into digital learning and teaching environments with the active participation of parents and students, ensuring participants’ wellbeing during e-learning, and using proven solutions are some of the main environmental factors that determine effective distance education (Kuzmanović et al., 2019). But there is another group of indicators within the individuals who introduce or should introduce new technologies to educational processes. In the case of active teachers, the situation is quite simple because this group is closely tied with the processes that take place in schools (Stošić & Stošić, 2015; Fedeli, 2017). Thus, many incorporate e-learning into organizational and methodology changes, and receive intense support in many forms. A group which seems less explored are pre-service teachers, that is, students of pedagogical specialisations. This group has limited experience as creators of e-learning. Such students are mainly users of distance education solutions (Istenič Starčič & Lebeničnik, 2020). Based on the existing research, it is believed that there are gaps in the system of preparing pre-service teachers. There are many studies into the existing generation of educators both working in K12 and higher education systems. Representing the generation raised in the information society as digital natives, the new teaching staff face diverse expectations. Stereotypically, it is this generation - current students - that is seen as possessing well-developed digital literacy and rich experience of using ICT. Such assumptions not only create expectations regarding the changes in the ways new media are implemented in education but also raise questions about the preparation of this next generation of teachers, who will face global challenges we can now observe thanks to global computerization, among other phenomena (Oyelere et al., 2020).

This paper joins a series of reports from research into the common points in pre-service teacher training, the development of e-learning, and the analyses of the determinants of the implementation of new media in academic and formal education. It presents the key determinants of using e-learning platforms, such as the individual educational experiences of students connected with e-learning, the self-evaluation of digital literacy level, the diagnosis of attitude towards new media, the evaluation of the quality of new technologies available at university, individual online activity, and ways in which mobile devices are used. According to specialists collaborating within an international project carried out in parts of Latin America, Europe and the Caribbean, the abovementioned variables are responsible for the effective implementation of e-learning solutions. They are crucial regardless of the geographical location of schools and universities.

2. Methodology

2.1 Research objective

The goal of the research was to show the relationship between the variables which determine the e-learning experiences of students of pedagogical studies. The dependent variable were experiences regarding e-learning during the last year, in particular participation in obligatory courses, searching for information connected with completing online classes, participation in paid and free online courses, and participation in self-learning groups held online. The above listed variables were compared with the following independent variables: attitude towards new media in education, subjectively perceived ease of use of ICT, quality of ICT available at the university, interest in online training, individual online activity, the means of use of mobile devices for different purposes, and self-evaluation of digital literacy. The following research questions are formulated in this paper: 1) What kind of experiences are characteristic of students of pedagogical faculties in the field of e-learning? 2) To what extent are the experiences of different forms of online learning linked? 3) Is participation in paid and free e-learning courses related to the attitude towards new media?

2.2 Research procedure and characteristics of the research sample

The research was conducted in the biggest state-owned Polish university that focuses on educating pre-service teachers - the Pedagogical University of Cracow. The study was carried out in May-June 2019 among the students of full-time, teacher studies in the Pedagogical Faculty. The research was part
of the international project Smart Ecosystem for Learning and Inclusion – SELI ERANet17/ICT-0076, financed by the National Centre for Research and Development (Tomczyk & Sunday Oyelere, 2019). The study was conducted in compliance with the appropriate ethical standards. The final version of the tool and the procedure were approved by researchers who represent nine countries participating in the SELI project. The respondents were informed about the research objectives and data processing procedures. Every respondent was able to withdraw from participating in the research at any time. The data collected were coded to prevent the identification of the persons completing the questionnaire. The study was conducted by researchers with many years of experience in social research. There were 450 respondents. The average age of the students was 22.660, with standard deviation of 4.232. The median was 21 years. The sample consisted of 395 (87.778%) women, 52 (11.556%) men and 3 individuals who did not declare their gender (0.667%). The data collected allow only for the generalization of research results for the population of students of the Pedagogical University of Cracow (Pedagogical Faculty).

2.3 Research tool

The tool was developed by a team of experts involved in the SELI project. The triangulation of approaches, experiences with e-learning, and the research sectors (IT and education) led to the final version of the tool which consisted of the following scales: attitude to new media (9 indicators) (Tomczyk et al., 2017); the perception of ease of use of ICT (4 indicators, description of SELI project), technical infrastructure at University – self-evaluation (7 indicators, description of SELI project), interest in future online courses offered in SELI (8 indicators, description of SELI project); use of the internet (including social networks and instant messaging apps) (9 indicators) (Eger et al., 2018); usage of internet for learning (5 indicators – original design); usage of mobile devices (5 indicators – original design); perception about the level of ICT skills (5 indicators) (Taubert, 2006; Petuhova et al., 2010). The whole tool was very coherent with Cronbach's $\alpha = 0.925$.

3. Results

The vast majority of the pre-service teachers interviewed have no regular experiences with e-learning. Most often, students of pedagogical specializations search for information online to prepare for their classes or find materials they can use in their projects (46.2% of the young people do this often of very often). More than one in four pre-service teachers (27.2%) have participated in obligatory online courses often of very often. However, these courses are most frequently health and safety, library orientation, or general open lectures (chosen by the students). Very rarely, distance courses are completed as part of methodology training. Until the COVID-19 pandemic, e-learning was used occasionally for courses preparing future teaching staff.

![Figure 1. Experiences of pre-service teachers with e-learning.](image-url)
One in five respondents (20.2%) has participated in free e-courses often or very often. The vast majority have never taken part in paid online training. Only 8.4% often or very often participate in courses that involve additional fees. Thus, e-learning is not a popular solution used by pre-service teachers to develop their interests or, more widely, to improve their human capital. As for ICT-based self-learning groups, the students also engaged very rarely. Only 15.1% had regular experiences in this area. Based on the data collected, one can notice that e-learning is not the leading means of acquiring academic knowledge or developing individual interests. It is rather a form or an environment for teaching and learning only occasionally used by pre-service teachers. The detailed distribution of answers is presented in Figure 1.

When analysing the data, it was found that participation in one form of distance education very often coexists with other forms of e-learning. Thus, individuals who actively acquire knowledge on the Internet try to benefit from the variety of opportunities provided in this process. This coexistence is confirmed in the analyses using Pearson’s linear correlation coefficient, the results of which are presented in Table 1. However, one must remember that the strength of these relationships is average. Thus, there must be some intermediate variables between the simple relationships, thanks to which the e-learning process is influenced by the series of other factors.

### Table 1. Correlations between different forms of remote education

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Study in an obligatory online course in my career or in my postgraduate studies</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Searching relevant sources on the Internet to complete online classes for my degree</td>
<td>0.604 ***</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Taking free e-learning courses (online courses - e.g. language, ICT)</td>
<td>0.477 ***</td>
<td>0.426 ***</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>4. Taking paid online courses</td>
<td>0.326 ***</td>
<td>0.235 ***</td>
<td>0.486 ***</td>
<td>-</td>
</tr>
<tr>
<td>5. Participating in online study groups</td>
<td>0.366 ***</td>
<td>0.308 ***</td>
<td>0.468 ***</td>
<td>0.579 ***</td>
</tr>
</tbody>
</table>

* p < .05, ** p < .01, *** p < .001

In order to illustrate the correlations between the variables that form the set of indicators of engagement in e-learning activities, individual and environmental indicators were included in the research model. Individually tested, each of these elements shows a positive linear correlation (due to volume limitations a detailed table of linear correlations between the indicators is not included here). However, given the total compilation of variables in the model with seven variables and the selected indicators of ICT-related experiences, it turns out that only one variable changes noticeably. This is the "use of the Internet" variable. The regularity involves both experiences with paid and free courses. The predictors in the research model adopted explain from several to a dozen or so percent (Table 2). Thus, these correlations do not allow us to draw clear conclusions about the individual and environmental factors that influence participation in e-learning. However, the adopted research model has statistical limitations, as it explains only 11.2% of cases (free online courses) and 6.2% (paid online courses).

### Table 2. Multilinear regression analysis - dependent variables: Taking free and paid e-learning courses

<table>
<thead>
<tr>
<th></th>
<th>Taking free e-learning courses (online courses - e.g. language, ICT)</th>
<th>Taking paid online courses (online courses - e.g. language, ICT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attitude to new media</td>
<td>0.084 0.053 -1.861 0.062</td>
<td>0.079 0.054 1.457 0.145</td>
</tr>
<tr>
<td>The perception of ease of use of ICT</td>
<td>-0.013 0.056 1.582 0.114</td>
<td>0.041 0.058 0.710 0.478</td>
</tr>
</tbody>
</table>

N=447
### Technical infrastructure at University

<table>
<thead>
<tr>
<th></th>
<th>-0.030</th>
<th>0.046</th>
<th>-0.231</th>
<th>0.817</th>
<th>0.011</th>
<th>0.047</th>
<th>0.250</th>
<th>0.801</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interest in future online courses</td>
<td>0.084</td>
<td>0.048</td>
<td>-0.657</td>
<td>0.511</td>
<td>0.012</td>
<td>0.049</td>
<td>0.258</td>
<td>0.796</td>
</tr>
<tr>
<td>Use of internet</td>
<td>0.146</td>
<td>0.057</td>
<td>1.740</td>
<td>0.082</td>
<td>0.145</td>
<td>0.058</td>
<td>2.470</td>
<td>0.013</td>
</tr>
<tr>
<td>Usage of mobile devices</td>
<td>0.087</td>
<td>0.060</td>
<td>2.560</td>
<td>0.010</td>
<td>-0.016</td>
<td>0.062</td>
<td>-0.266</td>
<td>0.789</td>
</tr>
<tr>
<td>Autoevaluation ICT skill</td>
<td>0.092</td>
<td>0.054</td>
<td>1.453</td>
<td>0.146</td>
<td>0.079</td>
<td>0.056</td>
<td>1.411</td>
<td>0.158</td>
</tr>
</tbody>
</table>

### 4. Discussion and summary

The data collected shed new light on the education of students of pedagogical studies in the biggest Polish university that prepares future teaching staff in modern schools. Of course, the sample does not enable generalisations of the whole student population in Poland but, as shown in other research, these data provide an illustration which may be useful for those who are responsible for the preparation of new teaching staff to function in the information society and enable the verification of the processes of implementation of obligatory academic courses like media in education or IT (Kędzierska & Wątek-Gozdek, 2015; Arteaga et al., 2020). One must be aware that the data are now historical due to the fact that since March 2020, all academic education-related processes in the university studied were moved into cyberspace. On the one hand, the data are historical but on the other hand they enable further, more complex and longitudinal research in the same sample.

During the research, the students of pedagogical specializations reported having had little experience with e-learning. The vast majority did not participate in paid online courses and only one in five used free forms of e-learning (Eger, 2015). Despite Poland being considered a country where information society services (banking, shopping, news services) are developing rapidly, e-learning is still not attractive for students. One question raised by the interpretation of the results is, what indicators block participation in e-learning? Based on the data collected, it was noticed that the issue is not connected with the respondents’ attitude to new media or to university infrastructure. The key indicator for different forms is the way in which digital services are used. Students who are active in cyberspace not only use typical services but benefit from the variety of opportunities provided by the Internet. Based on the results of the previous research, this group may be characterised as techno-optimists who recognise the power of media influence, very often experiment with the new media, and are aware that ICT may increase the quality of life (Wątróbki et al., 2018; Ziemba, 2017). The results indicate that financial wellbeing does not influence the frequency of participating in e-learning, and this is particularly interesting. At present, Polish students do not differ in this area from their peers in neighbouring countries (Eger et al., 2018). They use devices that perform well, and have smartphones with data packages that enable them to stay online anywhere and anytime. Thus, the indicators of the frequency of participation in e-learning should be sought in other areas, not necessarily related to the self-evaluation of digital literacy or attitude towards new media but rather in the means of using new media, which does not have to be connected with the respondents’ ability to use traditional software (for example office packages) (Tomczyk, 2019).

The results presented herein give grounds for further research into the individual and environmental factors that facilitate the development of digital literacy among the students of pedagogical specializations. Due to the changes taking place in the education system, it is the younger generations of teachers that will shape the image of schools in the next few years. Today, academic centres are particularly responsible for the effective preparation of this generation of teachers so that they can create online and traditional (offline) learning environments with methodological correctness and as a response to modern challenges, by combining the potential of both solutions.

### Acknowledgements

This work was supported by the ERANET-LAC project which has received funding from the European Union’s Seventh Framework Program. Project Smart Ecosystem for Learning and Inclusion – ERANet17/ICT-0076SELI. In Poland the project is financed by the National Centre for Research and Development (NCBiR). I would like to thank all of my colleagues involved in data collection, in
particular Dr Joanna Wnęc-Gozdek and Dr Anna Mróz. Separate thanks are due to the researchers involved in the SELI project who worked on the final version of the tool.

References


Tomczyk, L., & Sunday Oyelere, S. (2019). *ICT for Learning and Inclusion in Latin America and Europe. Case Study From Countries: Bolivia, Brazil, Cuba, Dominican Republic, Ecuador, Finland, Poland, Turkey, Uruguay*. Cracow: Pedagogical University of Cracow.


Supporting Group Learning Using Pen Stroke Data Analytics

Daichi YOSHITAKE a*, Brendan FLANAGAN b & Hiroaki OGATA b
aGraduate School of Informatics, Kyoto University, Japan
bAcademic Center for Computing and Media Studies, Kyoto University, Japan
*a@89e@gmail.com

Abstract: Education environments are increasingly becoming digitized, however in some circumstances handwriting input still is an important part of the learning process. In this paper, we propose and implement an educational support system using pen stroke data analysis. Junior high school teachers can perform data analysis themselves with the system. Pen stroke data is collected using a handwritten memo feature in the BookRoll digital learning material reader. Pen stroke data that is collected by the system includes: time, color, x-coordinate and y-coordinate. The system has functions for automatic group formation by using clustering based on the analysis of characteristics of learner pen strokes that are collected from the whole class. It is anticipated that the group formation function can support teachers in quickly creating learning groups in a timely manner based on automated data analysis.

Keywords: learning analytics, pen stroke data, handwriting, educational support, clustering, group learning, group formation

1. Introduction

In recent years, the introduction of information technology into education is progressing. However, there are few cases where learning log data is collected and analyzed to utilize when using digital teaching materials. As a field called learning analytics, research to analyze educational big data is also underway, but discussions have just begun (Hayashi & Hirashima, 2016). Learning analytics is a field that provides educational support by collecting and analyzing educational big data. For example, it is possible to grasp the status of the student’s learning efforts from the browsing history of the teaching materials, and recommend appropriate teaching materials.

Meanwhile, group learning is increasingly significant in education since not only cognitive knowledge, but also interpersonal skills such as critical thinking, problem solving and reasoning that count in modern society. Group formation is the most important component since it determines quality of group work (Wessner & Pfister, 2001). Group learning with appropriate groups is found to outperform traditional education (Kynndt et al., 2013). Teachers should create groups based on students’ performance and purposes of group learning. On the other hand, teachers usually take a long time on group formation, and it is difficult to create groups in a same class after students solved questions.

In this study, we focus on using pen stroke data obtained when students use a tablet and stylus pen to answer questions with a handwritten memo for group learning. By analyzing the pen stroke data, it is expected that a student’s stuck points while answering questions and their understanding level can be estimated. Pen stroke data analytics has already been studied for detecting answer stuck points (Iiyama et al., 2017). However, but group formation using pen stroke data analytics has not been studied. Since features such as stuck points are extracted from the course of the answer by handwriting analysis, it is possible to form a group from a viewpoint that is different from the previous research (Srba, I. & Bielikova, M., 2015). For example, using a heterogeneous group organization it increases the likelihood that students who have different stuck points could teach each other in the same group. In addition, pen stroke pressure is used in some research for educational support (Asai et al., 2012; Takahashi, Imoto, & Yamaguchi, 2015). However, if a system uses pen stroke pressure, students need to use a specified digital pen which supports pen pressure and thus limits the scale to which the analysis can be used. Therefore, in this study, we propose and develop a system that can provide educational support to students by analyzing pen stroke data, and the system is developed as a web application. This allows the system to not be dependent on what devices students use. We provide an interface that allows teachers to analyze pen stroke data themselves. If the data is analyzed by researchers, it cannot be
analyzed in real time, and teachers cannot provide immediate support students with results of data analytics. Teachers operate the system themselves to analyze data, so the usefulness of the system in an actual school can be confirmed.

The purposes of this study are to implement a system for group formation using pen stroke data analytics and to confirm the system is ready for being used in actual education sites. In this research, we proposed and implemented a system that allows teachers to analyze pen stroke data by themselves, and conduct a simulation of group formation. We anticipate that the group formation function can support teachers in quickly creating learning groups in a timely manner based on automated data analysis.

2. Digital Learning Material Reading System

2.1 BookRoll

Digitized learning materials are a core part of modern formal education. In addition to serving as a learning material distribution platform, it is also an important source of data for learning analytics into the reading habits of students. The action events of the readers are recorded, such as: turning to the next or previous page, jumping to different pages, memos, comments, bookmarks, and markers indicating parts of the learning materials that are hard to understand or are of importance. The reading behavior of students has previously been used to visualize class preparation and review patterns (Ogata et al., 2017; Yin et al., 2015). The digital textbook system can be used to not only log the actions of students reading reference materials, but also to distribute lecture slides.

In the present work, the non-proprietary BookRoll digital textbook system was used to serve lecture materials and capture learners reading behavior for analysis (Flanagan & Ogata, 2018). As shown in Figure 1, the user interface supports a variety of functions, such as: navigating to the next or previous page, jumping to any page, marking sections of reading materials in yellow to indicate sections that were not understood, or red for important sections. Memos can also be created at the page level or with a marker to attach it to a specific section of the page. Users can also bookmark pages or use the full text search function to find the information they need to review later. Currently, learning material content can be uploaded to BookRoll in PDF (Portable Document Format) format, and it supports a wide range of devices, including: laptops, tablets, and smartphones, as it can be accessed through a standard web browser.

![BookRoll interface](image)

*Figure 1. A screenshot of the BookRoll digital learning material.*

Reading behavior while using the BookRoll system is send using the xAPI standard for pseudonymized learning event logging and collected in an LRS (Learning Record Store). Table 1 presents a sample of BookRoll’s learner behavior logs that have been extracted from an LRS. In the logs there are many types of operations, for example, OPEN means that the student opened the e-book file and NEXT means that he or she clicked the next button to move to the subsequent page.
Table 1. A sample of events recorded from user interaction with BookRoll.

<table>
<thead>
<tr>
<th>Contents id</th>
<th>Memo text</th>
<th>Operation date</th>
<th>Operation name</th>
<th>Page no</th>
<th>User id</th>
</tr>
</thead>
<tbody>
<tr>
<td>EBOOK_341</td>
<td></td>
<td>2018/01/22 18:10</td>
<td>REGIST CONTENTS 0</td>
<td></td>
<td>T1</td>
</tr>
<tr>
<td>EBOOK_341</td>
<td></td>
<td>2018/01/23 9:10</td>
<td>OPEN</td>
<td>1</td>
<td>S1</td>
</tr>
<tr>
<td>EBOOK_341</td>
<td></td>
<td>2018/01/23 9:20</td>
<td>NEXT</td>
<td>2</td>
<td>S1</td>
</tr>
<tr>
<td>EBOOK_341</td>
<td>Sample memo</td>
<td>2018/01/23 9:21</td>
<td>OPEN</td>
<td>1</td>
<td>S2</td>
</tr>
<tr>
<td>EBOOK_341</td>
<td>Sample memo</td>
<td>2018/01/23 9:22</td>
<td>ADD MEMO</td>
<td>2</td>
<td>S1</td>
</tr>
<tr>
<td>EBOOK_341</td>
<td>Sample memo</td>
<td>2018/01/23 9:22</td>
<td>ADD HW MEMO</td>
<td>2</td>
<td>S1</td>
</tr>
</tbody>
</table>

2.2 Pen Stroke Input and Data

Figure 2 shows the BookRoll pen stroke input interface. Users can change the pen color and pen thickness, and use the eraser on the right-side menu to remove previously written pen strokes or correct mistakes. Pen stroke data is recorded automatically by BookRoll when a user writes strokes or uses the eraser function. This data is sent to the LRS for collection and analysis using the BookRoll standard xAPI interface as described in the previous section. The pen stroke data format is described below and an example can be found in the last row of Table 2 in the Memo Text column: Pen color, Multiple {UNIX time : x-coordinate : y-coordinate}. The pen color is recorded in rgb format. For example, if the pen color is black, the recorded data will be rgb(0, 0, 0). UNIX time is the time stamp when the data is record, and x-coordinate and y-coordinate express where the pen was at that particular time. UNIX time, and the x-coordinate and y-coordinate are multiple data and additional data is concatenated in a comma separated format, to make up a single pen stroke time series vector.

3. System

We will explain the configuration of the system proposed in this study. This system consists of BookRoll which is a teaching material distribution system, and Analysis Tool and an API for stroke analysis. Students use BookRoll’s handwriting memo function to answer questions. The Analysis Tool works with the API (Application Programming Interface) for stroke analysis. The API analyzes the sent stroke data and returns the result. We call this API Stroke API. Stroke API is made with Python and
Flask which is a Python framework. Flask is called a micro framework because it is light, and it works fast. It also takes only a few steps to build an API. 

First, students use BookRoll for writing handwritten memos. The pen stroke data from the memo are collected into the LRS automatically. Teachers then select the educational material which they want to analyze on Analysis Tool, and then Analysis Tool extracts the data from the LRS and sends it to the Stroke API. The Stroke API returns the analyzed results, and Analysis Tool then shows the result to the teachers. Teachers can give feedback to students and create learning groups based on the analytics.

Strokes are colored with red when the time interval is more than 2.5 seconds, orange in the case of more than 2.0 seconds, yellow for more than 1.5 seconds, green for more than 1.0 seconds, dark blue when it is more than 0.5 seconds, and black for less than 0.5 seconds. This feature allows teachers to check the student’s answer stuck points. Stroke API also supports the eraser function of BookRoll too. The output images from this API also include eraser stroke. In addition, Stroke API extracts the answer time from pen stroke data. Stroke API has a clustering function that clusters handwritten answers based on features using k-means. Given the number of clusters k, k-means classifies the data to be clustered into k clusters and calculates the average of each cluster. The following features are used for clustering: answer time, number of strokes and ratio of six colored strokes and eraser strokes. This system provides a way for teachers to understand student characteristics.

The system has a group formation function based on a result of clustering. The group formation function has 3 types of algorithms: homogeneous, heterogeneous and random algorithms. In the homogeneous algorithm, the system creates groups of students who have same cluster id because it can be said that students who have same cluster id have same features in their answers. In the heterogeneous algorithm, the system creates groups of students who have different cluster id. In the random algorithm, the system just creates groups randomly.

Steps of the homogeneous algorithm is below:
I. Sort students by a number of items in each cluster
II. Add students from the head of students list to a target group until a number of students of a group is satisfied
III. Change the target group to next group
IV. Repeat II and III until finishing formation

Steps of the heterogeneous algorithm is below:
I. Sort students by a number of items in each cluster
II. Add a student from the head of students list to a target group
III. Change the target group to next group
IV. Repeat II and III until finishing formation

Since groups are organized based on learning processes, it is expected to improve the quality of groups, which is important in group learning.

Figure 3 shows the user interface of the group formation function. Teachers set a number of students in 1 group, select a group formation algorithm, then the system creates groups automatically and show a result with student names and their cluster id. Numbers in brackets mean cluster id of those students. Teachers also can download a result as a CSV (Comma Separated Value) file.

![Figure 3. A screenshot of the group formation function.](image-url)
4. Simulation

We conducted a simulation of group formation using pen stroke data collected from a junior high school mathematics class in Japan. These pen stroke data were collected from 23 students by solving a question about quadratic equation with BookRoll’s handwriting memo function. A number of k-means clusters was set to 4. The system created groups based on results of clustering analysis and the number of students belong to a group was set to 4 or 5. We used the homogeneous algorithm and the heterogeneous algorithm to create groups in this simulation.

Table 2 shows the normalized average feature values of each cluster. Features of answer process are confirmed here. For example, students in cluster 2 took eraser much more than students in other clusters. Students in cluster 3 took a long time to write the answer.

Table 2. Normalized feature values of each clusters.

<table>
<thead>
<tr>
<th>Cluster id</th>
<th>Amount of stroke</th>
<th>Eraser</th>
<th>Black stroke</th>
<th>Blue stroke</th>
<th>Green stroke</th>
<th>Yellow stroke</th>
<th>Orange stroke</th>
<th>Red stroke</th>
<th>Answer time</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.28</td>
<td>0.12</td>
<td>0.94</td>
<td>0.90</td>
<td>1.00</td>
<td>1.00</td>
<td>0.73</td>
<td>0.13</td>
<td>0.19</td>
</tr>
<tr>
<td>1</td>
<td>1.00</td>
<td>1.00</td>
<td>0.41</td>
<td>0.42</td>
<td>0.43</td>
<td>0.26</td>
<td>0.85</td>
<td>0.34</td>
<td>0.57</td>
</tr>
<tr>
<td>2</td>
<td>0.40</td>
<td>0.05</td>
<td>0.95</td>
<td>1.00</td>
<td>0.93</td>
<td>0.38</td>
<td>1.00</td>
<td>1.00</td>
<td>0.41</td>
</tr>
<tr>
<td>3</td>
<td>0.58</td>
<td>0.05</td>
<td>1.00</td>
<td>0.53</td>
<td>0.76</td>
<td>0.33</td>
<td>0.74</td>
<td>0.70</td>
<td>1.00</td>
</tr>
</tbody>
</table>

5. Results

Table 3. A result of group formation using the homogeneous algorithm.

<table>
<thead>
<tr>
<th>Group 0</th>
<th>Student in</th>
<th>Student in</th>
<th>Student in</th>
<th>Student in</th>
<th>Student in</th>
<th>Student in</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
</tr>
<tr>
<td>Group 2</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
</tr>
<tr>
<td>Group 3</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
</tr>
<tr>
<td>cluster[0]</td>
<td>cluster[0]</td>
<td>cluster[0]</td>
<td>cluster[0]</td>
<td>cluster[0]</td>
<td>cluster[0]</td>
<td>cluster[0]</td>
</tr>
<tr>
<td>Group 4</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
</tr>
</tbody>
</table>

Table 4. A result of group formation using the heterogeneous algorithm.

<table>
<thead>
<tr>
<th>Group 0</th>
<th>Student in</th>
<th>Student in</th>
<th>Student in</th>
<th>Student in</th>
<th>Student in</th>
<th>Student in</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
</tr>
<tr>
<td>Group 2</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
</tr>
<tr>
<td>Group 3</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
</tr>
<tr>
<td>Group 4</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
<td>Student in</td>
</tr>
</tbody>
</table>

Table 3 and Table 4 show the results of the group formation simulation. The number in the brackets are the cluster id. We acquired groups of some students who have similar or different types of answers. In Table 3, Groups 0, 2, and 3 were created with only students with the same cluster id. Group 4 has one student who has a different cluster id, but this is inevitable because there is only one student with cluster id 0. There is also only one student with cluster id 1 in group 1. It is possible to place more than one student with the same cluster id if the system exchange students with another group. In this simulation, distances between each cluster were not assumed. If the system considers the distances between clusters, groups would have some students who have different cluster ids, but they still have similar features.

In Table 4, about the heterogenous algorithm, if the number of students in one cluster is large for the number of students with a certain cluster id, a group with all has different ids cannot be created. There is possibility that the heterogeneous algorithm is improved by considering the distance between
clusters like the homogeneous algorithm.

6. Conclusions

We proposed and implemented a system for group formation based on a clustering result using pen stroke data. The system creates heterogeneous, homogeneous or random groups. We conducted a simulation of group formation with pen stroke data that was collected from a junior high school mathematics class in Japan. In this paper, we confirmed the system functions as intended by conducting simulation and currently preparations are being made to examine the effectiveness of the system though a formal evaluation experiment in a real classroom.

As future work, we are going to conduct a demonstration experiment to evaluate the system. First students solve mathematics questions using BookRoll’s handwriting function. Then, a teacher chooses the heterogeneous or homogeneous algorithm for a purpose and creates groups of students using the system with the chosen algorithm and the random algorithm. We are going to compare performance and feedback of students of groups created by the chosen algorithm and the random algorithm.

Furthermore, considering distances between each cluster for both of the homogeneous algorithm and the heterogeneous algorithm to improve results of group formation including different or same cluster ids.

Acknowledgements

This work was partly supported by JSPS Grant-in-Aid for Scientific Research (B)20H01722, JSPS Grant-in-Aid for Scientific Research (S)16H06304 and NEDO Special Innovation Program on AI and Big Data 18102059-0 and SPIRITS 2020 of Kyoto University.

References

What Influence Malaysian Teachers’ Satisfaction towards the FROG Virtual Learning Environment? A Structural Equation Modelling

Mei Lick CHEOK\textsuperscript{a}, Su Luan WONG\textsuperscript{b}\textsuperscript{*}, Ahmad Fauzi MOHD AYUB\textsuperscript{c} & Rosnaini Mahmud\textsuperscript{d}

\textsuperscript{a}Form Six College, Melaka, Malaysia
\textsuperscript{bcd}Universiti Putra Malaysia, Malaysia
\textsuperscript{*suluan@upm.edu.my}

Abstract: The purpose of this paper is to determine teachers’ perceptions towards factors that influence their satisfaction towards a learning management system in Malaysia. A survey which involved 350 teachers from three states and the focus of the study was to investigate factors that influence satisfaction towards the FROG Virtual Learning Environment (VLE) in the teaching and learning among the Champion Secondary Schools teachers. The factors were computer anxiety, computer attitude and internet self-efficacy, training, technical support, school management perceive, usefulness, perceived ease of use, flexibility and interaction in studying FROG VLE’s characteristics. The study was based on a quantitative method with correlational research design. The study involved 350 respondents who were selected using the proportionate stratified cluster sampling. There were four significant paths in influencing satisfaction which are the internet self-efficacy, computer attitude, training, and flexibility.

Keywords: FROG VLE, satisfaction, teachers, Malaysia

1. Introduction

The Malaysian Education Policy Review conducted by UNESCO (2013) reported Malaysia as lagging behind in terms of technology integration in education in comparison to many other countries in the region. Improving and empowering teachers and school leadership was given top priority under the Malaysian Education Blueprint 2013-2025 (MEB). Educators including school principals need to realise the importance of equipping themselves with new technologies in the 21\textsuperscript{st} learning environments (Afshari, Bakar, Luan & Siraj, 2012). MEB is a detailed plan of actions that maps out the education landscape for the next 13 years (2013-2025). Under the MEB, 1BestariNet Project was initiated in order to leverage on the Internet and technology use to bridge the digital divide between the rural and urban schools. A single virtual learning platform known as the FROG Virtual Learning Environment (FROG VLE), and a high-speed 4G connectivity were provided to all the schools in Malaysia. This virtual learning platform linked the six million school children from 10,000 schools in the area of 329, 847 km with 4.5 million parents and teachers (Hew & Kadir, 2016).

2. Literature Review

Implementation of any virtual learning environment (VLE) can be expensive to any organisations due to the relatively low adoption rate among users. This is because there is a low widespread change in pedagogic practices despite the varied functionality afforded by the VLE (Becker & Jokivirta, 2007). Concerns are being raised regarding the economic cost of implementing and maintaining the infrastructure in order to sustain the integration of technology into the classrooms (Dutta & Bilbao-Osorio, 2012). Given the increasing availability and reliance of technology in the modern world, there is a dire need to understand factors that can help lead us to sustain and increase its adoption (Yee, Luan, Ayub & Mahmud, 2009). Reasons as to why it worked or failed to work need to be understood.
Numerous past research have revealed that satisfaction is among the most important factor in the success of system implementation (Martin-Rodriguez & Fernandez-Molina, 2014) for reason being that it ensures continued usage of the system (Joo & Choi, 2016). For this reason, it would make sense that when teachers are satisfied with the FROG VLE, they will continue using it even after its initial implementation. Teachers shared vision and commitment for the initial uptake of innovative learning technology and continuation of e-learning initiatives in schools are critically needed.

2.1 Rationale of the Study

Despite the massive expenditure and enthusiasm by the MOE, a report by the Auditor-General’s (A-G) report (National Audit Department, 2014) revealed that usage of the FROG VLE by teachers, students and parents was between 0.01 and 4.69 percent while daily utilisation of the VLE by teachers was found to be between 0.01 and 0.03 percent. This seems to suggest that the VLE is underused or unused by most of the teachers.

As past studies have found that satisfaction does influence teachers’ continuation of web-based learning system usage (Al-Busaidi & Al-Shihi, 2012; Sun, Tsai, Finger, Chen & Yeh, 2008) and has been used as a dependent variable in e-learning research (Teo, 2014; Teo & Wong, 2013), it must also be considered in our local context. At the same time, variables like computer attitude (Yu & Yang, 2006), computer self-efficacy (Chen, Yeh, Lou & Lin, 2013), computer anxiety (Ozkan & Koseler, 2009), perceived usefulness (Teo, 2014), perceived ease of use (Teo, 2014), interaction (Martin-Rodriguez, Molina, Alonso & Gomez, 2014), flexibility (Ho, Nakamori, Ho & Lim, 2016), management support (Ho, Nakamori, Ho & Lim, 2016), training (Aggelidis & Chatzoglou, 2012), and technical support (Ozkan & Koseler, 2009) have been found to influence satisfaction towards a learning management system. Past researchers have revealed that satisfaction is among the most important factors in the success of system implementation and it is influenced by the different facets of user satisfaction that can be attributed to various dimensions: teachers’ factors, system design and environmental factors (Wang & Bagakas, 2003). Given the high stakes in e-learning and the growing reliance on technologies in education, there is a dire need for a research to be done in Malaysia to probe the determinants of satisfaction that would entice teachers to accept and continue to use FROG VLE in their teaching and learning processes. Only when teachers are using the FROG VLE can we expect a change in the teaching and learning environment.

2.2 Objective and Hypotheses of the Study

Given the aforesaid context and the dearth of information on studies related to FROG VLE in Malaysian schools, it is therefore, pertinent that a study on factors that influence satisfaction towards the FROG VLE be conducted in Malaysian schools. The independent variables used in this study are based on previous studies which include perceived usefulness, perceived ease of use, interaction, flexibility, computer attitude, computer anxiety, internet self-efficacy, technical support, training and school management. The following hypotheses were formulated as shown in Table 2.

<table>
<thead>
<tr>
<th>Hypotheses of the study</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1: Computer Attitude has a significant influence on teachers’ satisfaction towards the FROG VLE in teaching and learning</td>
</tr>
<tr>
<td>H2: Internet Self-Efficacy has a significant influence on teachers’ satisfaction towards the FROG VLE in teaching and learning</td>
</tr>
<tr>
<td>H3: Computer Anxiety has a significant influence on teachers’ satisfaction towards the FROG VLE in teaching and learning</td>
</tr>
<tr>
<td>H4: Perceived Usefulness has a significant influence on teachers’ satisfaction towards the FROG VLE in teaching and learning</td>
</tr>
<tr>
<td>H5: Perceived Ease of Use has a significant influence on teachers’ satisfaction towards the FROG VLE in teaching and learning</td>
</tr>
<tr>
<td>H6: Interaction has a significant influence on teachers’ satisfaction towards the FROG VLE in teaching and learning</td>
</tr>
<tr>
<td>H7: Flexibility has a significant influence on teachers’ satisfaction towards the FROG VLE in teaching and learning</td>
</tr>
<tr>
<td>H8: School management support has a significant influence on teachers’ satisfaction towards the FROG VLE in teaching and learning</td>
</tr>
<tr>
<td>H9: Training has a significant influence on teachers’ satisfaction towards the FROG VLE in teaching and learning</td>
</tr>
<tr>
<td>H10: Technical support has a significant influence on teachers’ satisfaction towards the FROG VLE in teaching and learning</td>
</tr>
</tbody>
</table>
3. Methodology

3.1 Sampling and Instrumentation

The population in this study are teachers from the 27 Champion FROG VLE secondary schools located in the southern region from the states of Negeri Sembilan, Melaka and Johor. The FROG VLE Champion schools are schools listed in Malaysia as the benchmarked schools, acting as mentors to the others (Hew & Kadir, 2016). These schools were selected so as to ensure that all the teachers involved in the study are using the FROG VLE in their classrooms. There are 155 secondary schools in the whole of Malaysia which are listed in the Champion list. The sample consists of 350 teachers; 25 males and 60 female teachers in Negeri Sembilan, 24 males and 60 females in Melaka while 49 males and 132 female teachers from Johor which are proportionally represented according to the population studied.

The questionnaire comprised 132 items on a 5-point Likert scale with 1 representing “strongly disagree” and 5 representing “strongly agree” for positive items (and vice versa for negative items). Table 3 shows the internal consistency of the instrument. Cronbach’s alpha coefficient was utilised to determine reliability of the questionnaire. The Cronbach’s alpha ranged from .98 to .92 which revealed that the internal consistency estimation appeared adequate and above the cutoff value of .70 (Nunnally, 1978).

<table>
<thead>
<tr>
<th>Scale</th>
<th>Cronbach’s alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived Usefulness</td>
<td>.98</td>
</tr>
<tr>
<td>Flexibility</td>
<td>.97</td>
</tr>
<tr>
<td>Perceived Ease of Use</td>
<td>.96</td>
</tr>
<tr>
<td>Interaction</td>
<td>.96</td>
</tr>
<tr>
<td>Computer Attitude</td>
<td>.92</td>
</tr>
<tr>
<td>Internet Self-Efficacy</td>
<td>.96</td>
</tr>
<tr>
<td>Computer Anxiety</td>
<td>.97</td>
</tr>
<tr>
<td>School Management</td>
<td>.92</td>
</tr>
<tr>
<td>Technical Support</td>
<td>.95</td>
</tr>
<tr>
<td>Training</td>
<td>.96</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>.98</td>
</tr>
</tbody>
</table>

4. Model Fit

Structural equation modeling (SEM) was performed to test the fit between the research model and the obtained data through AMOS software Version 22. SEM is suitable for theory testing and confirmation where there are strong theoretical foundations and hypotheses that are driven by model fitness. Table 3 shows the level of acceptable fit and the fit indices for the proposed research model. In the case of $x^2$, it has been found to be sensitive to sample size differences, especially for cases in which the sample size exceeds 200. As such, this anomaly is assumed to be applicable in the present study with a sample of 350. However, the results of the other fit indices shows a good fit for the proposed research model.

<table>
<thead>
<tr>
<th>Fit Index</th>
<th>Recommended Level of Fit</th>
<th>Proposed Research Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x^2$</td>
<td>P value &gt; .05</td>
<td>1678.276</td>
</tr>
<tr>
<td>$x^2$/df</td>
<td>&lt; 5</td>
<td>1.9</td>
</tr>
<tr>
<td>GFI</td>
<td>&gt; 0.90</td>
<td>.826</td>
</tr>
<tr>
<td>CFI</td>
<td>&gt; 0.90</td>
<td>.956</td>
</tr>
<tr>
<td>NFI</td>
<td>&gt; 0.90</td>
<td>.912</td>
</tr>
<tr>
<td>TLI</td>
<td>&gt; 0.90</td>
<td>.950</td>
</tr>
<tr>
<td>RMSEA</td>
<td>0.05 ≤ RMSEA ≤ 0.10</td>
<td>.051</td>
</tr>
</tbody>
</table>

Figure 1 shows the resulting path coefficients of the proposed research model. Four out of the ten hypotheses were supported by the data. The results showed:
1. significant influence of computer attitude on teachers’ satisfaction towards the Frog VLE ($\beta = 0.08, p < 0.05$);
2. significant influence of internet self-efficacy on teachers’ satisfaction towards the FROG VLE ($\beta = 0.11, p < 0.05$);
3. no significant influence of computer anxiety on teachers’ satisfaction towards the FROG VLE ($\beta = -0.04, p > 0.05$);
4. no significant influence of perceived usefulness on teachers’ satisfaction towards the FROG VLE ($\beta = 0.13, p > 0.05$);
5. no significant influence of perceived ease of use on teachers’ satisfaction towards the FROG VLE ($\beta = -0.04, p > 0.05$);
6. no significant influence of interaction on teachers’ satisfaction towards the FROG VLE ($\beta = 0.03, p > 0.05$);
7. significant influence of flexibility on teachers’ satisfaction towards the FROG VLE ($\beta = 0.16, p < 0.05$);
8. no significant influence of school management on teachers’ satisfaction towards the FROG VLE ($\beta = -0.11, p > 0.05$);
9. significant influence on training on teachers’ satisfaction towards the FROG VLE ($\beta = 0.48, p < 0.05$);
10. no significant influence of technical support on teachers’ satisfaction towards the Frog VLE ($\beta = 0.02, p > 0.05$).

Figure 1: Path Coefficients of the proposed research model

5. Discussion

The resulting model is an adequate fit to the investigated factors that are believed to influence satisfaction since 86% of the variance in satisfaction was explained by the four constructs (computer attitude, internet self-efficacy, flexibility and training).

Training was found to have a significant effect on satisfaction. This suggests that with training, teachers are more likely to be satisfied towards the FROG VLE. This finding is similar to those of Khasawneh and Yaseen (2017) who found a strong correlation between training and satisfaction. The results also suggest that internet self-efficacy has a significant influence on satisfaction. This implies that with more confidence and believe towards their own capabilities in handling usage of working on the internet, teachers will be more satisfied towards the FROG VLE. Flexibility is also significant in this study. Arbaugh (2002) found flexibility as one of the main factors that influence learner satisfaction.
in an online learning environment. Results from the analysis also seem to suggest that Computer Attitudes has an influence on satisfaction towards Frog VLE. This is in line with the findings by Sun, Tsai, Finger, Chen and Yeh (2008) who found that it is an important element in predicting satisfaction among e-learner volunteers. Perceived usefulness does not have a direct effect on teachers’ satisfaction towards the FROG VLE. This contradicts the studies by Stockless (2018), and Chen (2010) who found that perceived usefulness influenced their respondents’ satisfaction. Perceived ease of use is found to be insignificant in this study. This contradicts many previous studies which suggest otherwise like in Arbaugh and Duray (2002) and Wu, Hsia, Liao and Tennyson (2008). The results also indicated that Interaction was not significant to influence satisfaction towards the FROG VLE. This is in direct contrast to the findings by Ramayah and Lee (2012) who found interaction to be an antecedent of satisfaction towards online instructional courses. Computer Anxiety was not found to have a significant effect on satisfaction. With the many ICT projects by the MOE over the past nearly thirty years, they have enabled our teachers to be competent in at least the basic ICT skills, thus leading them to overcome their anxiety. Lastly, school management and technical support are not significant in the analysis. This contradicts Thurmond, Wambach, Connors and Frey (2002) who in their studies claimed that environmental factors were highly predictive on whether or not users were satisfied with the web-based lessons.

6. Conclusion

The challenge in the ICT scenario in Malaysian education system is to get teachers to want to use FROG VLE considering the large amount of money spent. As this study seemed to suggest, stakeholders especially the MOE need to pay more attention to teachers’ Internet Self-efficacy, Computer Attitude, Training and Flexibility as these have been found to play more important roles in influencing teachers’ satisfaction.

References


International Student Perspectives on Simulated Preservice Induction

Jon MASON*, Nicolas GROMIK©, Khalid KHAN© & Bing LIU©

©College of Education, Charles Darwin University, Australia
*jon.mason@cdu.edu.au

Abstract: Responding to an expanding cohort of international pre-service student teachers, the College of Education at Charles Darwin University instigated a pilot study into the viability of SimLabTM as a platform for induction into Australian classroom dynamics prior to placement. SimLab is a mixed reality platform that can be configured to simulate a variety of real-life experiences that preservice teachers on placement will likely encounter. In recent years, it has been used successfully by several Australian universities as an embedded component within initial teacher education courses. A series of 15-minute sessions were provided to three cohorts of international students (n=52). Based on a simple 5E pedagogical model that prioritizes engagement as a foundation for all teaching, students were tasked with getting to know their students and to deliver at least one micro-teaching session. Additionally, they were provided with opportunities to have conversations with the school principal and their assigned mentor teacher. Initial findings suggest a favourable overall evaluation. The College is now extending this research by providing this opportunity to all incoming preservice teachers.

Keywords: SimLab, simulation, teacher education, preservice, international, mixed reality

1. Introduction

Providing Pre-Service Teachers (PSTs) with sufficient exposure to pre-employment classroom experiences is a constant challenge for universities and colleges of education. While PSTs may complete the required theory-based learning, they may not necessarily transfer this knowledge into applied teaching practices. To assist them, various engagement strategies have been implemented such as facilitating peer-feedback (Benton-Kupper, 2001) or enabling them to reflect on their video recorded micro-teaching performance (Önal, 2019). Effective scaffolding has also been reported through enabling students to ask questions based on their interactions with their own performance data using smart technologies (Khan, & Mason, 2017, 2018).

The research literature on utilising technology for supporting PST training also identifies micro-teaching as a viable pedagogical approach, particularly within technology enabled contexts (Arsal, 2015; Ping, 2013). While micro-teaching now has broader connotations (Scagnoli, 2012), we refer to Cruickshank and Metcalf’s (1993) definition in which it is presented as a way for PSTs to practice teaching and demonstrate their teaching capabilities, strengths and weaknesses for a small class for a short time. Thus, using video-based micro-teaching, Kennedy and Lees (2016) report on a project that encouraged participants to receive feedback from coaches. Key outcomes were that participants “reported few technical difficulties in filming” and developed “a clearer understanding of their perceived strengths and weaknesses” (p.377) (see also Black et al., 2016).

With technological advancements such as digital classroom simulation, encouraging students to participate in video-recorded micro-teaching seems easier. For international students, it also provides a less threatening option for initial exposure to a classroom than real-life. After identifying the availability of trialing SimLab for an extended period without too much cost, this research project was therefore developed to offer international students experience in teaching practice within a simulated Australian classroom context as part of their preparation.

As a “controlled learning environment” that reduces real classroom challenges, SimLab enables students to acquire some foundation skills and aptitude to manage a classroom (Ledger & Fischetti, 2020, p. 37). In related research, Black et al. (2016) and Hudson, Voytecki and Zhang (2018), report that digital classroom simulation environments provide participants with experience to a selected risk
and stress free, context. With the capability to configure a range of scenarios to interact with, SimLab is managed and operated by a real academic that follows a script whilst playing various roles of selected characters in an avatar shape. Scenarios used in our project varied from meeting the principal, teaching in a classroom, and meeting a mentor (see Figures 1 to 3 below). Brought to life by real-life interactors, the characters can raise their hands or move on their chairs, although they cannot stand up or move beyond their desks. It is also possible for a supervisor to monitor interactions within this environment.

![Figure 1. Meet the Principal](image1)  ![Figure 2. Meet the Students](image2)  ![Figure 3. Meet the Mentor](image3)

Images courtesy of Murdoch University with permission

The PST interacts with the class with limited assistance for about fifteen minutes. Initially, this is focused on becoming acquainted with the digital student characters. This is followed by an opportunity to deliver a lesson. The primary objective of this pilot study was to assess whether international PSTs would find the experience beneficial. A secondary objective was to enable post-experience observation of the interactions for both students and researchers. As Dieker et al., (2014) report, simulation-based teaching environments offer students with a digital interactive classroom setting to practice their understanding of various teaching strategies and approaches in a risk-free context. This research, then, was designed as a pilot study of SimLab to determine whether it is an effective induction experience for international PSTs into Australian classrooms.

2. Methodology

This research used mixed methods involving triangulation of data collection. All participants first completed a pre-test survey informed by the Technology Acceptance Model. Observation of PSTs of video-recorded SimLab sessions provided a rich source of interaction material. Participants then engaged in a short interview following the sessions and completed a post-test survey. It also draws on the 5E pedagogical model originally developed by Bybee and Landes (1990) and used extensively within the Australian schooling system. This model places emphasis on engagement as the crucial first step in teaching. Related research (Ledger & Fischetti, 2020) clearly demonstrates the effectiveness of this platform for PSTs; this project, however, was focused only on international PSTs.

2.1 Participants

All international PSTs located on campus within an initial teacher education course offered by the College of Education were invited to experience the SimLab classroom environment to prepare for their professional placement. Fifty-two students signed the consent form and agreed to participate in the study. A teaching schedule was organized, and students were invited to trial the SimLab environment first through meeting the ‘students’ and to follow this by delivering a lesson of their choosing. They were also asked to complete a pre-test survey. The survey showed most students (n=40) participants had no prior professional experience while about one fifth (n =11) students indicating, having already experienced their first placement in their current teaching course. With regards to having competence with using technology for learning, participants reported strongly agreeing (n=31) moderately agreeing (n=10) and somewhat agreeing (n=3) with the statement. Some students chose not to answer these. The volunteering participants seemed to have no bias or inclination towards the SimLab.
2.2 Research design

In designing the overall PST experience, we chose to designate a purpose-built room for this research. Because all participants were based on campus it seemed easy to do this, although students still had to physically go to this venue rather than sitting with their own computer in a familiar setting. Doing this thereby added another dimension to the simulated experience – that is, the students did not know precisely what they would first encounter. The room selected included a large television monitor fitted with speakers, a webcam with a microphone, all connected to a computer (see Figure 6 & 7 below). The sessions were organized via Zoom and the performance recorded on the computer.

An invitation to participate was circulated to all international Pre-Service Teachers located on campus. This was followed by a simple sequence of steps with each participant:

- Completion of a pre-test survey to determine levels of technology familiarity and competence
- An initial 15-minute session focused on gaining familiarity with the SimLab environment
- Delivering a 15-minute lesson in SimLab classroom.
- Meeting with the Principal and/or Mentor teacher prior to experiencing a disruptive session.
- A short interview about the session was conducted.
- A post-test survey was completed.

As mentioned, research design included pre- and post-surveys, with a SimLab experience in between. The pre-survey collected demographic background information about the students prior learning experiences. The SimLab experience consisted of volunteers either meeting and/or delivering a lesson to a group of five digital avatar-based classroom students, for a period up to 15 minutes, which also included a brief introduction to SimLab, and an exit discussion about their digital experience. Finally, at the start of the following semester, participants were invited to answer a post-survey to gauge their recollections and opinions about their SimLab experience. Both pre- and post-surveys were conducted online using Qualtrics.

Nine statements were included by participants in the post-test survey in addition to a Likert scale to gauge levels of agreements or non-agreement:

1. I had a very good experience.
2. The SimLab session resembles a real classroom experience.
3. I get to know better about Australian classroom after the session.
4. I know better of the importance of classroom management through this experience.
5. I know better of the importance of lesson planning through this experience.
6. The training in SimLab is an effective way to practice classroom skills.
7. I am better prepared for my placement or to teach lessons after my SimLab session.
8. After SimLab sessions, I am more confident that I can engage students in the future.
9. If there are more SimLab sessions in the future. I am willing to attend again.

3. Data Collection and analysis

The pre-test survey collected information about student demographics. For the SimLab sessions, observations were made based upon the performance of students in this environment. Following their SimLab sessions, a short interview was conducted with focus on the student’s experience of SimLab. Each interview was short, lasting about 2 minutes. In addition, a summary of their performance was
recorded. The typical five-level Likert items were used (definitely yes, probably yes, might or might not, probably not, and definitely not) to classify responses. The percentage of each item is retrieved from the Qualtrics report. The qualitative data were coded following thematic analysis. The quantitative data from Qualtrics report were used to complement the qualitative findings.

4. Summary of Findings

The evidence provides some critical insights into the SimLab experience from the point of view of a Pre-Service Teacher and in line with research objective. Much of it matches expectations, particularly after talking with students directly after their experience.

As a summary evidence, students found the experience appealing and engaging and believed they gained benefit from doing so. For the researchers, observing the video recordings confirmed this – there was significant diversity in the style of interaction while most students achieved good rapport within the simulated classroom. Following sections describe these findings in greater detail.

4.1 Engagement opportunities

In total, 52 PSTs participated in the study and took part in pre-survey. Of these, 28 participants answered the post-test survey. Data from this survey indicated that a significant majority of PSTs felt better prepared for placement after their SimLab experience.

To the question regarding their experience of SimLab, 64.29% (18 out of 28) participants responded definitely yes and 32.14% (9 out of 28) participants chose probably yes.

Participants indicated that through SimLab they gained a better understanding of the importance of classroom management with 39.29% responding definitely yes, and 35.71% indicated probably yes. 21.43% were somewhat ambivalent, responding might or might not. 3.57% indicated probably not.

On better understanding of importance of lesson planning through SimLab 42.86% indicated definitely yes, 39.29% showed probably yes, 14.29% indicated might or might not. Only 1 participant indicated definitely not.

Following their experience with SimLab, 32.14% of participants strongly agreed that they are more confident in being able to engage with classroom students in the future, 39.29% indicated that they somewhat agree. 17.86% neither agree nor disagree with that.

18 participants of 28 (64.29%) clearly indicated that they would like to attend SimLab in future. 25% responded probably yes to this. However, 10.71% indicated that they will probably not attend SimLab in the future.

4.2 Genuine interaction

Though most these PSTs failed to provide a well-prepared lesson for students in SimLab, the interactions between the PSTs and SimLab students were genuine. SimLab provided great flexibility with AI and interactor. The student avatars in SimLab responded naturally and spontaneously to the PST participants. Participants later stated they did not expect the student avatars from SimLab were real people.

Participants indicated from the post survey that SimLab session resembled a real classroom experience. According to the post survey, 32.14% of participants strongly agree and 57.14% somewhat agree that the SimLab session resembled a real classroom experience.

The data on participants reaching a high agreement on SimLab informing them about Australian curriculum remained inconclusive, with 38.46% of participant indicated definitely yes. 26.92% indicated probably yes, 15.38% think might or might not. 19.23% showed probably not.

For the question of whether training in SimLab provides an effective method to practice classroom skills, participants had diverse opinions. With 42.86% indicating strong agreement and 35.71% showed somewhat in agreement. However, 14.29% showed that they neither agree nor disagree. 7.14% participants strongly disagree with the idea.

As for whether SimLab better prepared student for their placement or to teaching, 32.14% expressed they strongly agree and 32.14% indicated somewhat agree with that. Yet 28.57% of
participant neither agree nor disagree. About 7.14% participants strongly disagreed that SimLab could help with their placement or teaching.

The post-survey also included an open-ended question engaging respondents to offer their opinion about the experience. In terms of SimLab providing participants helpful experience, to prepare with understanding the Australian School Systems, its policies and procedures in general, three respondents provided the following responses.

Case 1
I remembered 5 of the students that I have connected with. They were all different from one another. They were from different backgrounds and cultures. The variety gave me a choice to talk about various topics hence it gave me the opportunity to initiate ideas to include diversity.

Case 2
I only spent 15 minutes with the students, but my most vivid memory was the challenge I had to interact with simulated representations of real people. Beforehand I thought it would be easy since most people like myself are used to interacting with computers and playing computer games. However, the experience was extremely intimidating. I believe that over time, one could get used to this. But in general, I thought it was a great opportunity, even if it was only to check my own preconceptions.

Case 3
It was extremely difficult to not look past the fact that in the sim, you were working with visual representations and not the real thing. Similar to not confusing a character in a 3rd Person type computer game with a real human being.

Overall, the outcome of this pilot experiment was of benefit to volunteer participants with the evidence indicating that international pre-service teachers were enthusiastic about experiencing and exploring a virtual classroom. It provided them with some confidence prior to undertaking their professional experience placement. In addition, session times of 15 minutes exposure to SimLab was enough for students to experience and explore topics or teaching attempts – importantly, it enabled them to focus on the situation at hand. Although the interactions were with avatars, the experience was dynamic and immersive which provided a sense of real engagement.

5. Limitations and Further Research

5.1 Limitations

Most SimLab interactions indicated that participants gained familiarity with the environment and their engagement with the avatars was through relying on general topics such as hobbies, favorite sports, international festival, and food. However, we also wanted them to focus on the important first step of the 5E pedagogical framework which is all about engagement. Given the funding constraints, this limited further exploration of the platform in terms of other aspects of the 5E framework.

There was a significant attrition rate between pre- and post- sessions surveys. The main reason for this is that many of the students felt they needed to compare the real-life experience of professional placement to be able to properly respond to the post-test survey, which they lacked at the time of this study. Despite this, however, the total number of participants were within the project protocol.

Another limitation concerned the equipment used. The microphone used was not of very high quality and some conversations were not well recorded.

5.2 Further Research

The evidence from this study can be considered as informative and as a first step to ongoing, more structured research framed by the 5E pedagogical framework. We aim to extend the SimLab experience to all students (domestic and international) enrolled in teacher education. Moreover, given the social distancing constraints that immediately followed this study due to the pandemic, we envisage a greater role with simulation technology in the future.
6. Conclusion

As the preliminary process and evidence suggests, testing SimLab as a digital simulation classroom environment, was successful for both students and researchers. While the infrastructure and approach may at times not have been the most effective, improvements to provide a full immersive experience and enhance the data collection process can be adapted. The evidence suggests that the volunteering participants appreciated the opportunity to meet with digital students and interact in a risk-free environment. SimLab is a professionally engaging simulation platform that international students were able to engage in, with and through, and this warrant further research in a growing industry.

Acknowledgements

We would like to acknowledge the College of Education at Charles Darwin University for providing a small grant (H19065) to enable this study.

References


Development of Computational Thinking Concepts in Scratch Programming

Siu-Cheung KONG\textsuperscript{a*}, Ming LAI\textsuperscript{b} & Cora Ka-Yuk SIU\textsuperscript{b}

\textsuperscript{a}Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong

\textsuperscript{b}Centre for Learning, Teaching and Technology, The Education University of Hong Kong, Hong Kong

\texttt{*sckong@eduhk.hk}

Abstract: Young learners can use block-based programming environments to develop their computational thinking (CT), but as previous studies suggest, students’ CT concepts development should not be taken for granted. This paper presents a study of the development of CT concepts of repetition, conditionals, and sequences in senior primary school students through an interest-driven curriculum in a four-day Scratch programming summer camp. The participants were 30 students, 17 boys and 13 girls, who had just finished grade 4. A mixed methods design was used, with quantitative data to examine the effectiveness of the curriculum in developing students’ CT concepts and qualitative data to explore the students’ learning when creating their final projects. Based on the pre- and post-test results of a CT concepts test, significant improvements were found for each of the concepts, with a particularly large effect size on repetition. This was in line with the design of the curriculum, which placed slightly more emphasis on repetition. An analysis of the students’ final projects indicated that the girls’ groups were more likely to choose the theme of storytelling over creating a card. In general, the students tended to use the concepts and features that were taught more in the curriculum, and were able to use them in a creative way in their final projects.

Keywords: Computational thinking concepts, Concept development, Primary school students, Project outcomes, Scratch programming

1. Introduction

The development of students’ digital creativity is an important educational goal in the 21st century. Pioneers such as Seymour Papert laid the foundations for this goal along with the seminal paper by Wing, who described that computational thinking (CT) should be an essential skill for everyone (Wing, 2006). CT has since been incorporated into K-12 education worldwide (Tang, Chou, & Tsai, 2019), and the development of visual block-based programming environments such as Scratch has allowed even young learners to develop CT by using these environments (Brennan & Resnick, 2012). Although promising results have been achieved in students’ CT development, empirical studies have revealed that students can still encounter difficulties in learning CT concepts when using these block-based environments (Zhang & Nouri, 2019). This paper presents a study of the development of senior primary school students’ CT concepts through an interest-driven curriculum with a focus on the core CT concepts of repetition, conditionals, and sequences, in a four-day Scratch programming camp.

2. Literature Review

CT involves “solving problems, designing systems, and understanding human behavior, by drawing on the concepts fundamental to computer science” (Wing, 2006, p.33). Brennan and Resnick (2012) proposed a CT framework consisting of three dimensions: 1) the CT concepts students engaged with in programming; 2) the CT practices developed through engaging with the concepts; and 3) the perspectives formed about the world and oneself through CT activities. As block-based programming environments have become increasingly popular in recent years, studies have been conducted into students’ learning of CT concepts, as reviewed in the following section.
2.1 CT concepts development when using block-based programming environments

Among the various CT concepts that can be found in the literature, sequences, repetition (loops), and conditionals were identified as the basics for young learners (Zhang & Nouri, 2019). Franklin et al. (2017) found that a high percentage of grade four to grade six students were able to finish tasks related to sequences, and concluded that the concept was accessible to students at grade four or even earlier. The concept of repetition may be more difficult for students (Grover & Basu, 2017), and they may not know the conditions and procedures required to terminate a loop (Grover, Pea, & Cooper, 2015), but some students were found to improve their understanding of repetition after learning in Scratch (Meerbaum-Salant, Armoni, & Ben-Ari, 2013). Meerbaum-Salant et al. (2013) found that students were able to grasp the concept of conditionals, but they may not fully understand the concept, particularly when the conditional block was combined with other blocks, such as nested within a forever loop (Lye & Koh, 2018). These studies suggest that it cannot be assumed that students’ understanding of CT concepts will necessarily improve when using a block-based programming environment.

2.2 Research questions

In this study, we examined the effectiveness of an interest-driven curriculum in developing students’ understanding of key CT concepts. The research questions were: 1) Did primary students’ CT concepts improve significantly after attending a four-day Scratch programming summer camp? How did they apply the concepts they had learned in the summer camp in the final projects they created?

3. The Curriculum

The curriculum was designed to develop the CT competency of Primary 4 students and encourage their digital creativity. Based on Brennan and Resnick’s (2012) CT framework, it involves interest-driven learning activities that trigger, immerse, and extend their interest in computational thinking and foster their learning habits (Kong, 2016) using Scratch programming. The Interest-driven Creator (IDC) theory indicates that it is important to arouse learners’ interest while increasing their computational thinking capabilities and enforcing learning habits (Kong, 2016). Thus, developing CT concepts with guided learning resources and creative tasks that encourage their creativity is essential. These learning habits can be further extended through the primary 5 and 6 curricula, which are not covered in this paper. This curriculum consists of seven units and a final project, as shown in Figure 1.

| 1. Creative Computing | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 2. Dancing Cat | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 3. Make a Maze Game | ✓ | ✓ | ✓ | ✓ | ✓ |
| 4. Tell a Joke | ✓ | ✓ | ✓ | ✓ |
| 5. Tell a Story | ✓ | ✓ | ✓ | ✓ |
| 6. Make Magic | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 7. Computational Arts | ✓ | ✓ | ✓ | ✓ |
| Final Project | Game | Digital Narrative | Digital Animation | With Physical Computing | Sequences | Conditionals | Repetition |

Figure 1. Computational thinking curriculum for Primary 4 students

This curriculum was designed to engage students through step-by-step guided activities, including creating games, stories, animations, and integrating physical computing objects for arousing students’ interests (Farris & Sengupta, 2016). After completing the seven units, the students paired up to design and implement their own projects using the CT concepts they had learned. They could either create a card or tell a story. This enabled them to apply their CT capabilities and digital creativity.

Sequences, repetition, and conditionals are the three core CT concepts taught in the curriculum, as shown in Figure 1. Four of the seven units in the curriculum involved repetition, and this was reinforced through authentic learning experiences to consolidate the learners’ understanding. In particular, the “Computational Arts” unit required learners to draw a square by duplicating the codes for drawing a line and then turning 90 degrees to the right four times. By asking the students whether a
pattern could be identified and what could help to simplify the task, they were able to see the actual value of repetition. **Sequences** relate to the order of code blocks. In Scratch, “broadcasting” can be regarded as a type of sequences that addresses the time-dependent issue, particularly in projects that involve digital narratives and animations, such as storytelling. The units of “Tell a Joke” and “Tell a Story” allowed students to understand the CT concept of sequences, and how they could articulate a story smoothly and sequentially. **Conditionals** were involved in the two game-making units “Make a Maze Game” and “Make Magic.” Here, interactions with users were enabled based on algorithms specified as conditions. The interest of the learners was aroused through playing the games and was further extended as they became immersed in coding and debugging.

### 4. Methodology

#### 4.1 Research Method

A mixed methods design was implemented, with quantitative data to examine the development of students’ CT concepts, and qualitative data to assess their learning when creating projects.

#### 4.2 Participants and the summer camp

The participants were 30 primary students, 17 boys and 13 girls, who had just finished grade 4 at the same school. Their average age was 10. Over the four days, they were taught by an instructor who was involved in the development of the curriculum. Their school teacher was in the classroom but was not involved in the teaching. The seven units were taught in the first two days and the morning of the third day, and the students completed their projects over the third and fourth days. They then presented their projects. In completing the projects, students worked in pairs, which were assigned by their school teacher. The total instructional time, including the students’ presentations, was around 16 hours.

#### 4.3 CT concepts test

A multiple-choice CT concepts test was developed to assess student learning in the summer camp. This involved three categories of CT concepts: 1) “repetition”, running the same sequence several times; 2) “conditionals”, decision-making based on conditions; and 3) “sequences”, identifying the steps of a task. As mentioned, these are all important concepts in the design of the curriculum. The test contained 14 items, with 4 “repetition”, 6 “conditionals”, and 4 “sequences” items respectively. For example, in a conditionals item, students were presented with codes with an if-then block, and were asked to select what would happen when the codes were executed. The test lasted for 30 minutes. The overall consistency of the test, as indicated by the Cronbach alpha in the post-test of this study, was 0.70.

#### 4.4 Students’ projects and classroom observation

The projects created by students were first analyzed by examining the choice of project themes, and second, by examining the blocks in the Scratch projects to determine the CT concepts applied. Besides, during the project presentation of students, a researcher sat at the back of the classroom and took field notes as a form of non-participant observation. The notes consisted of the description of the events and the thoughts of the researcher (Creswell, 2015). The notes helped us understand how the students applied the concepts they had learned and their digital creativity in completing the projects.

#### 4.5 Procedures

The students answered the CT concepts test at the beginning of the first day of the camp (pre-test). Then at the end of the fourth day, they answered the test again (post-test). Each time they had 30 minutes to complete the test. They were told to finish the test alone and no discussion was allowed.
5. Results

5.1 CT concepts

We averaged the correctness of all items to represent an overall score for the CT concepts. We also averaged the correctness of items of the categories of repetition, conditionals, and sequences respectively. Table 1 gives the results of the paired sample t-tests, which indicated significant improvements for the CT concepts test as a whole and for individual concepts. To examine the effect size of each comparison, we calculated the Cohen’s d, with a value of .20, .50, and .80 indicating a small, medium, and large effect size respectively. The results in Table 1 suggest a large effect for repetition and a medium effect for conditionals and sequences.

Table 1. Paired t-tests of CT concepts with effect sizes

<table>
<thead>
<tr>
<th>CT concepts</th>
<th>Pre-test</th>
<th>Post-test</th>
<th>t-value</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Repetition</td>
<td>.63</td>
<td>.31</td>
<td>.83</td>
<td>.27</td>
</tr>
<tr>
<td>Conditionals</td>
<td>.66</td>
<td>.27</td>
<td>.73</td>
<td>.28</td>
</tr>
<tr>
<td>Sequences</td>
<td>.59</td>
<td>.15</td>
<td>.68</td>
<td>.21</td>
</tr>
<tr>
<td>Overall</td>
<td>.63</td>
<td>.18</td>
<td>.75</td>
<td>.19</td>
</tr>
</tbody>
</table>

*: p<.05; **: p<.01; ***: p<.001

To examine whether there was any gender difference in the learning gains of the students in terms of CT concepts, an Analysis of Covariance (ANCOVA) on the post-test scores, with the corresponding pre-test scores as covariates, was conducted. The gender differences were not significant for any of the concepts, implying that both boys and girls gained similarly in terms of each CT concept.

5.2 Analysis of student projects

The project groups consisted of six groups of girls, eight of boys, and one mixed group. Projects under the theme of “birthday card” could further be differentiated into those applying features learned in computational arts and those applying other Scratch features. Four of the six girls’ groups chose storytelling in their projects, as did the mixed group. One girls’ group chose birthday card with computational arts and the other chose birthday card with other Scratch features. The choice of theme of the boys’ groups was more diverse, with three choosing storytelling, five choosing birthday card with computational arts, and two choosing birthday card with other Scratch features. The choice of theme of the boys’ groups was more diverse, with three choosing storytelling, five choosing birthday card with computational arts, and two choosing birthday card with other Scratch features. Table 2 presents how the CT concepts or Scratch features appeared in the projects. “Repetition” appeared in all projects of birthday card using computational arts. The use of “conditionals” was not that frequent in any project category. “Broadcast” as a form of “sequences” appeared in all projects with computational arts, two thirds of the birthday card projects using other Scratch features, and half of the storytelling projects.

Table 2. Percentage of projects with the appearance of concepts

<table>
<thead>
<tr>
<th>CT concepts/Scratch features</th>
<th>Storytelling projects (n=8)</th>
<th>Birthday card using computational arts (n=4)</th>
<th>Birthday card with other Scratch features (n=3)</th>
<th>All projects (n=15)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repetition</td>
<td>37.5%</td>
<td>100%</td>
<td>33.3%</td>
<td>53.3%</td>
</tr>
<tr>
<td>Conditionals</td>
<td>12.5%</td>
<td>25%</td>
<td>0%</td>
<td>13.3%</td>
</tr>
<tr>
<td>Broadcast</td>
<td>50%</td>
<td>100%</td>
<td>66.7%</td>
<td>66.7%</td>
</tr>
</tbody>
</table>

5.3 Student learning in projects

Students’ learning could also be indicated in the classroom observations of their project presentations. One of the boys’ groups created a project under the birthday card theme by using...
what they had learned in the computational arts unit. Their program began with the drawing of two snowflake patterns simultaneously, one at the top and the other at the bottom of the screen, as shown in the left of Figure 2. After drawing the patterns across the screen, the letters “HAPPY BIRTHDAY” appeared, as shown in the right of Figure 2.

![Figure 2. Birthday card by drawing two snowflake patterns at the top and bottom simultaneously](image)

In their presentation, the students mentioned that the most difficult part was to synchronize the drawing of the two patterns. As Figure 2 shows, the upper pattern consisted of squares, which took less time to draw than the pattern at the bottom consisting of circles. The students addressed this problem by intentionally manipulating the number of loops, creating redundant squares at the top so that the drawing times required for both top and bottom patterns became similar. Thus, although only 12 squares were needed to draw the snowflake pattern at the top, by trial and error they increased this to 32 squares for synchronizing the drawing time. This suggested that they were able to apply what they had learned about the concept of repetition and use it creatively to address their specific problem.

6. Discussion

In this paper, the development of students’ understanding of CT concepts is examined through an interest-driven curriculum in a four-day Scratch programming camp. The results of the CT concepts test indicated significant improvement in the understanding of all the concepts assessed. A large effect size was found for the concept of repetition, which was in line with the curriculum design in which students were exposed to the concept of repetition in several units. The concept of conditionals was covered in two of the units, significant improvement was still found, although the effect size was not as large as that of repetition. The appearance of CT concepts and Scratch features in the students’ final projects was also in line with the design of the curriculum, as repetition and broadcast had a higher percentage of appearance than conditionals. These results imply that the application and development of CT concepts are strongly affected by the curriculum design, which supports the literature suggesting that the development of students’ understanding of CT concepts should not be taken for granted (Zhang & Nouri, 2019), and the design of the curriculum is critical.

When examining the choice of project themes, it was found that the girls’ groups were more likely to choose storytelling. Other studies have suggested that storytelling could be a possible strategy for enhancing girls’ interest in CT activities (Burke & Kafai, 2012). However, the difference in gender preference for project themes did not appear to affect the students’ conceptual development, as there was no gender difference in the improvement in the understanding of any of the CT concepts. Thus, although storytelling projects may not involve the concept of repetition as much as projects with computational arts (see Table 2), the girls in this study still improved their understanding of the concept of repetition as much as the boys did. Thus, storytelling could be embedded in the curricular design to enhance the learning interest of girls, and such a design would not limit their development of the understanding of CT concepts.

Developing digital creativity is an important educational goal in the 21st century. The case presented in this study suggests that students were able to apply the CT concepts they had learned, and even use them creatively in their projects. They manipulated the number of loops to create redundant squares so that the drawing time could be synchronized, but they had not been taught this in the curriculum. Thus, in solving the synchronization problem they faced, they applied the concept of repetition creatively. The results suggest that with sufficient time to explore and create their own final
projects and with the solid foundation of CT concepts developed in previous units, students are able to apply their CT knowledge and exercise digital creativity.

6.1 Limitations

One limitation of this study was that it involved a pre-post design, with no control group included for comparison, hence the development of students’ CT concepts might not be due to the curriculum but other factors. However, the t-test results suggested that the students improved more on the concept of repetition than conditionals, which was in line with the curriculum design. Moreover, as the setting was in a four-day summer camp, the results might not be generalizable to other contexts. Hence further studies are needed to see whether the curriculum is effective in an ordinary classroom context.

Acknowledgements

The author would like to acknowledge the support of the coolthink@JC project funded by the Hong Kong Jockey Club Charities Trust.

References


Learning Support System for Understanding Pointers Based on Pair of Program Visualizations and Classroom Practices

Koichi YAMASHITA*\textsuperscript{a}, Keisuke SAKATA\textsuperscript{b}, Satoru KOGURE\textsuperscript{b}, Yasuhiro NOGUCHI\textsuperscript{b}, Tatsuhiro KONISHI\textsuperscript{b} & Yukihiro ITOH\textsuperscript{c}

\textsuperscript{a}Faculty of Business Administration, Tokoha University, Japan
\textsuperscript{b}Faculty of Informatics, Shizuoka University, Japan
\textsuperscript{c}Shizuoka University, Japan
*\textsf{yamasita@hm.tokoha-u.ac.jp}

Abstract: In this paper, we describe a program visualization (PV) system with capability of visualizing logical data structures and concrete memory space in the execution process of a program based on teachers’ intent of instruction, and classroom practices for understanding pointers using this system. Thus far, we have developed a PV system called TEDViT which has two distinctive features: capability of customizing PV of target domain world based on teachers’ intent of instruction, and capability of simultaneous visualization in target domain world and memory image which have different levels of abstraction. Here, target domain world is a notional world of data processed by the target programs, visualized as logical data structures. However, TEDViT evenly visualizes a memory space as memory image, thus, providing insufficient support for novice learners to understand program behaviors by observing and comparing target domain world and memory image visualizations. Therefore, in this study, we extended the original TEDViT to enable teachers define highlighting expressions in PV of memory image so that they could designate focus points or provide natural language descriptions to learners. We evaluate the effect of learning support that our extended system provided to novice learners by introducing the system in actual classrooms. The evaluation results based on the score improvements between pre and post-tests suggest that our extended TEDViT and classroom practices would have a certain degree of learning effectiveness.

Keywords: Programming education, program visualization system, program visualization design, pointer learning

1. Introduction

To this end, we have developed a program visualization (PV) system called TEDViT (Teacher’s Explaining Design Visualization Tool), and have conducted several classroom practices using this system (Kogure et al., 2014). There are two distinctive features of TEDViT: first, it makes it possible for teachers to customize PVs based on their intent of instruction, and second, it makes it possible for learners to observe and compare two PVs with different abstraction levels of target data structures. However, from our own experience of several classroom practices, we consider that the visualizations provided by TEDViT are not sufficient to help novice learners understand pointers. For instance, for target domain world, teachers can customize PVs and designate focus points by highlighting the changes made in the domain world by the target program by providing comments with balloon, etc. Meanwhile, the visualization of memory image provides almost the same level of information as a debugger since TEDViT can only flatly visualize the contents of memory space, and cannot express teachers’ intent of instruction by highlighting the changes. If the learners are software engineers who have an intuitive grasp of the so-called notional machine (Sorva, 2013), they may be able to perceive the change in the memory image and detect points to focus based on their own experience. However, novice learners with insufficient background knowledge might not perceive the changes.

Therefore, in this study, we extended TEDViT to make it possible for teachers to define highlighting expressions for target domain world as well as memory image. The PV customizations for
target domain world are achieved by defining drawing rules in TEDViT. The additions include a rule system to define highlighting expressions for memory image and a function to visualize memory image. We conducted two classroom practices for understanding pointers using our extended TEDViT, incorporating it into actual programming courses for novices. The evaluation results based on pre- and post-tests suggest that the extended TEDViT would have a certain degree of effectiveness in novice learners’ understanding of pointers.

2. Previous Works

TEDViT allows teachers to define the policy for drawing a status of the target domain world based on their intent of instruction. Teachers can create or edit a configuration file independently from the target program file. TEDViT interprets such a visualization policy by scanning the configuration file and then visualizes the target domain world accordingly. The learners can then observe the program behavior in the target world visualized in accordance with the teacher’s intent of instruction. At the moment, TEDViT supports visualizations of C programs only.

Figure 1 shows a learning environment generated by TEDViT. The environment comprises three fields: the data structures processed by the program in (A) are visualized in (B) and (C). TEDViT reproduces a series of memory images of variables in (B) for each step of the program’s execution, and a series of statuses of the target domain world in (C) that visualizes logical data structures. When a learner clicks the “Next” or “Prev” button, the highlight in (A) moves to the next or previous statement in the program code, the memory image in (B) is updated according to the values of the variables after executing the highlighted statement, and the corresponding status of the target domain world is visualized in (C). TEDViT simulates statement execution step by step so that the learner can understand the program’s behavior by observing the changes in the target domain world in (C). The learner can also understand the concrete memory image in each execution step and the concrete expression of the data structures by observing and comparing the target domain world in (C) with the memory image in (B).

However, compared with the visualizations of the target domain world in (C), which include highlighting expressions and descriptive comments defined by teachers, the flat visualizations of the memory image in (B), which lack customizability, make it difficult to perceive changes in values and understand what is happening in memory space. For this reason, we have almost never conducted classroom practices introducing TEDViT with learning scenario including observations of PVs in (B). Although an evaluation result suggested that observations of PVs in (B) would have some effectiveness in a practiced class targeting software engineers (Yamashita et al., 2017), we consider that it was because the subjects had some expectations about the changes in memory image from their experiences...
and had observed PVs in (B) based on these expectations. However, for novice learners, TEDViT provides inadequate visualizations in (B) to support their understanding of program behaviors.

3. Extending Visualizations of Memory Image Field

We extended the drawing function and rule system of TEDViT to visualize drawing objects in memory image field based on teachers’ intent of instruction. The insufficient support for novices in the original TEDViT had arisen from difficulties in perceiving changes in values and intuitively grasping the focus points in PVs in the memory image. Therefore, we included additional drawing objects in the extended TEDViT to visualize highlighting expressions. In addition, we analyzed several PVs from various teaching materials and textbooks, and decided to introduce the following extensions to TEDViT.

Ext1. Function to highlight the value of a specific pointer variable and the area pointed to by the pointer in memory image field.
Ext2. Function to set character color, background color, font size, and framed box of specific value in memory image field.
Ext3. Function to draw a framed box around the area corresponding to a specific variable in memory image field.
Ext4. Function to draw a text box of descriptive comment in memory image field.

Figure 2 shows a memory image visualization generated using the abovementioned extensions. Ext1 is a function to draw a framed box around the value of a specified pointer variable, an arrow toward the area pointed to by the pointer, and a framed box around that area. A single drawing rule is required to define Ext1, specifying the target pointer variable and the number of area blocks pointed to by that target pointer. Ext2 is a function to highlight a specified single value in memory image field. The single value can be that of a variable, or the address of a variable, variable name, and type of variable. Ext3 is a function to set a framed box around the entire area visualizing information of a specified variable. Ext4 is a function to display a message in memory image field, specifying drawing position as the position relative to the specified single value.

The memory image field visualizes memory area corresponding to the processing data of target program in tabular form. The drawing unit is a variable, an array element, or a structure member, each of which is visualized in a single column. The displayed item in each row corresponds to a prescribed attribute, such as the address of the data, reference name in the target program, and its value. It is easy to specify a particular column and row, hence how to draw positions or highlight targets can be specified in drawing rules without any ambiguity. As mentioned earlier, the drawing rule in the extended TEDViT consists of a condition and an object part. While the former can be described in the same way as was done in the original TEDViT, the latter can be specified using the abovementioned extensions. We consider that drawing rules in the extended rule system can be defined without any confusion with some experience in rule definitions.

4. Evaluation
To evaluate the effectiveness of learning scenario for novice learners including observations of PVs in memory images and the learning support of the environment generated by the extended TEDViT, we conducted classroom practices introducing the system in two actual programming courses for different students of the same university. We evaluated system effectiveness based on pre- and post-test scores. Table 1 provides a brief summary of the classroom practices.

Table 1. Summary of Classroom Practices

<table>
<thead>
<tr>
<th>Class</th>
<th>#1</th>
<th>#2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject (N)</td>
<td>90</td>
<td>40</td>
</tr>
<tr>
<td>Subjects’ major</td>
<td>Computer Science</td>
<td>Mathematical and System Engineering</td>
</tr>
<tr>
<td>Subjects’ grade</td>
<td>Sophomore</td>
<td>Freshman</td>
</tr>
<tr>
<td>Course</td>
<td>Algorithms and Data Structures I</td>
<td>Fundamental programming</td>
</tr>
</tbody>
</table>

Class #1 was incorporated into the course “Algorithms and Data Structures I” for sophomores majoring in Computer Science. The subjects had one-year experience in Java programming but had no experience in C. The aim of the class was to understand pointers in connection with a memory model as a C specific feature not found in Java, and had the following three goals.

G1-1. To understand address and pointer operators (i.e., reference and dereference operators).
G1-2. To understand subscript operators and relationships between subscript and pointer operators.
G1-3. To understand the implementation and working of pointers to structure variable.

In the Class, the teacher who regularly taught the course lectured on fundamental syntax of C language and pointers for 90 minutes, and then conducted a pre-test for 10 minutes. The pre-test consisted of three questions corresponding to the three goals, presenting the subjects with three programs similar to the ones discussed by the teacher in the lecture and asking about related concepts.

Q1-1. Asking the meaning of the values obtained by address and pointer operators at certain points in the execution process.
Q1-2. Asking about the reference target when an array was accessed without subscript operators.
Q1-3. Asking the meaning of the value obtained by address operation with a structure variable.

After the pre-test, the teacher allowed students to use our extended TEDViT to observe program behaviors without giving them any feedback on the test. The target was three sample programs which the teacher discussed in the lecture session. The teacher explained how to launch and operate the learning environment generated by the extended TEDViT and gave the students 10 minutes each to observe the three target program behaviors. During this observation session, the teacher did not provide any explanation about the target program behaviors but directed students to observe and compare the three fields of TEDViT according to visualizations on the learning environment. The drawing rules of the extended TEDViT included visualizations in the target domain world and highlighting expressions in memory image, and was defined by the teacher beforehand based on his own intent of instruction.

Following the observation session, which lasted 30 minutes, the teacher conducted a post-test for 10 minutes. The content of the post-test was almost same as the pre-test, except the assigned values of some variables and constant values in some arithmetic expressions. Table 2 shows the average rate of correct answers for pre- and post-tests in the class. The average rate for each question significantly improved after introducing the extended TEDViT, suggesting that this classroom practice had a certain degree of effectiveness.

Table 2. Average Rates of Correct Answers for Pre- and Post-Tests in Class #1

<table>
<thead>
<tr>
<th>Question</th>
<th>Pre-test</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1-1</td>
<td>.562</td>
<td>.844</td>
</tr>
<tr>
<td>Q1-2</td>
<td>.236</td>
<td>.692</td>
</tr>
<tr>
<td>Q1-3</td>
<td>.382</td>
<td>.788</td>
</tr>
</tbody>
</table>

Class #2 was incorporated into the course “Fundamental Programming” for freshmen majoring in Mathematical and System Engineering, as the second class in two consecutive lessons on pointers.
The subjects had less than a year of programming experience. The aim of the class resembled that of Class #1, and included the following goals.

G2-1. To understand the fundamental concepts of pointers such as address and pointer operators, pointer increment and decrement, and so on.
G2-2. To understand behaviors of “pseudo” call by reference by using pointers in function call.
G2-3. To understand subscript operators and relationships between subscript and pointer operators.

The overall flow of Class #2 was almost the same as that of Class #1. A teacher who regularly taught the course lectured the students for 90 minutes on pointers, coding exercises for three example programs corresponding to the three goals, and operation verifications of sample programs. After that, the teacher conducted a pre-test for 10 minutes which consisted of three questions corresponding to the three goals.

Q2-1. Asking the values obtained by expressions including pointer and increment operators at certain points in the execution process.
Q2-2. Asking the return values of function call with and without pointer variables.
Q2-3. Asking the values of array elements assigned using several assignment expressions including subscript, pointer, and increment operators.

After the pre-test, the teacher allowed students to use our extended TEDViT to observe program behaviors without giving them any feedback on the test. The target was the same three example programs discussed in the coding exercises. The teacher explained how to launch and operate the TEDViT learning environment and gave the students 15 minutes each to observe the three target program behaviors. As in Class #1, here too the teacher did not provide any explanation about the program behaviors and defined the drawing rules beforehand based on his own intent of instruction.

Following the observation session, which lasted 45 minutes, the teacher conducted a post-test for 10 minutes. The content of the post-test was almost same as the pre-test, except the assigned values of some variables and constant values in the provided programs were changed from the pre-test. Table 3 shows the average rate of correct answers for pre- and post-tests. The average rate for each question significantly improved after introducing the extended TEDViT, suggesting that this classroom practice had a certain degree of effectiveness. The reason behind the small improvement in average rate for Q2-3 compared to the other questions could be the insufficient understanding of arrays. The subjects in Class #2 did not have much programming experience and had learned arrays just before the lessons on pointers. The similar reason also holds for the relatively low rate of correct answers for Q2-3 in both pre- and post-tests.

Table 3. Average Rates of Correct Answers for Pre- and Post-Tests in Class #2

<table>
<thead>
<tr>
<th>Question</th>
<th>Pre-test</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q2-1</td>
<td>.256</td>
<td>.581</td>
</tr>
<tr>
<td>Q2-2</td>
<td>.535</td>
<td>.814</td>
</tr>
<tr>
<td>Q2-3</td>
<td>.165</td>
<td>.279</td>
</tr>
</tbody>
</table>

In addition, to evaluate the cost of defining the highlighting expressions in memory image field, we measured the time required to define the drawing rules for two programs $p_1$ and $p_2$ consisting of about 30 statements. The number of subjects of this pilot experiment were eight: two teachers who have experience of teaching programming, two students who have experience as programming teaching assistants, and four students who have the same level of programming experience as teaching assistants. First, we explained the drawing rules and rule system of the extended TEDViT to the subjects, and discussed an example program consisting of about 20 statements and drawing rules consisting of highlighting expressions in memory image field for the example program. Then, the subjects observed another PV without any drawing rules and were asked to define the drawing rules that reproduce the provided PV. The time taken by each subject to define the drawing rules was measured. To reduce the order effect, we divided the subjects into two groups, each of which included one teacher, one teaching assistant, and two students, and swapped the order of target program for rule definitions. The number of defined rules by each subject was about 20 to 30, although there were some differences according to the subject. Table 4 details the result of this pilot experiment. Every subject took less time to define rules for the second target program than the first one, suggesting that experience could improve the
productivity of rule definitions. The average time to define highlighting expressions in memory image field for the second target program is 26.05 minutes.

Table 4. Time Taken to Define PVs by each Subject

<table>
<thead>
<tr>
<th>Subject #</th>
<th>1st target</th>
<th>Time</th>
<th>2nd target</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$p_1$</td>
<td>40m 25s</td>
<td>$p_2$</td>
<td>37m 01s</td>
</tr>
<tr>
<td>2</td>
<td>$p_1$</td>
<td>25m 30s</td>
<td>$p_2$</td>
<td>21m 22s</td>
</tr>
<tr>
<td>3</td>
<td>$p_1$</td>
<td>50m 03s</td>
<td>$p_2$</td>
<td>30m 59s</td>
</tr>
<tr>
<td>4</td>
<td>$p_1$</td>
<td>34m 19s</td>
<td>$p_2$</td>
<td>26m 55s</td>
</tr>
<tr>
<td>5</td>
<td>$p_2$</td>
<td>23m 06s</td>
<td>$p_1$</td>
<td>21m 00s</td>
</tr>
<tr>
<td>6</td>
<td>$p_2$</td>
<td>30m 24s</td>
<td>$p_1$</td>
<td>20m 38s</td>
</tr>
<tr>
<td>7</td>
<td>$p_2$</td>
<td>31m 15s</td>
<td>$p_1$</td>
<td>24m 40s</td>
</tr>
<tr>
<td>8</td>
<td>$p_2$</td>
<td>45m 12s</td>
<td>$p_1$</td>
<td>25m 48s</td>
</tr>
</tbody>
</table>

5. Conclusion

In this paper, we describe a PV system with capability of defining highlighting expressions for the target domain world as well as memory image, and two classroom practices introducing the system in actual classes. We extended the rule system of TEDViT and added the functions for rule interpretation and drawing memory image field. This extension allowed teachers to define highlighting expressions using the extension, and it was expected that learners could perceive changes in values of memory spaces and detect points to focus. To evaluate the effectiveness of the learning support provided by the extended TEDViT and the learning scenario, we conducted classroom practices introducing it in actual classes. As the learning target, we focused on pointers which many novice learners struggle with. The evaluation results based on pre- and post-test scores clarified that the learners cultivated a better understanding of pointers, hence, suggesting that our classroom practices introducing the extended TEDViT would have a certain degree of effectiveness. Furthermore, to evaluate the cost of classroom preparation using our extended TEDViT, we measured the time taken by eight subjects including actual teachers to define drawing rules of some highlighting expressions in memory image field. This evaluation result suggests that teachers can define PVs based on their intent of instruction in actual classes using the extended rule system which includes visualizations of highlighting expressions in memory image field. Therefore, we conclude that the extended TEDViT in this study makes it possible to define PVs based on teachers’ intent of instruction, and that it would have a certain degree of effectiveness in helping novice learners understand pointers.

Acknowledgements

This study was supported by JSPS KAKENHI Grant Numbers JP18K11566, JP18K11567, and JP19K12259.

References


663
Distance Learning Practices: A Reflective Study

Ali AHMED*, Craig WATTERSONa, Nilufar BAGHAEIb, Karsten LUNDQVISTa

a School of Engineering and Computer Science, Victoria University of Wellington, New Zealand
b School of Natural & Computational Sciences, Massey University, Auckland, New Zealand

*ali.ahmed@vuw.ac.nz

Abstract: Well-planned online delivery of content can be a challenging experience and is of significant importance, especially these days due to COVID-19 pandemic. Online delivery of computer security subjects in a pure distance learning environment is no exception. The modules should have high quality and be designed in an engaging, interactive and fun way for students, who happen to have different motivation levels and educational backgrounds. The quality of education for distance learning degrees varies significantly, so does the experience of the teaching staff. There have been few studies focusing on the teaching practices of online degrees especially those with a computer security major. This work presents a comprehensive reflective paper that demonstrates the evolution of a world-class distance learning degree at the University of Liverpool (UoL) in the UK over the last ten years. We report on the teaching practices adopted in cybersecurity-related modules/courses as well as the experience of the first author (who is a subject matter expert for cybersecurity-related modules) in teaching those modules. The paper describes the structure of the modules and how online teaching/learning is taking place along with the assessment types. Besides, we discuss both positive and negative observations and the lessons learnt. We believe our work can pave the way for successful and well-planned online delivery of cybersecurity modules in future.

Keywords: Distance Learning, Online Students, Teaching Cyber-security, Assessment Rubrics, Student Satisfaction

1. Introduction

Distance learning has gained a high momentum during the current COVID-19 pandemic, where education providers across the globe have been forced to cancel all face to face classes and move their courses online, in most cases with limited notice or preparation time. While some institutions have both the experience and resources to handle this situation smoothly, many others have been struggling to convert the traditional course materials to suit online learners and provide meaningful learning experiences. The lack of preparation time was also a major factor that affected the online conversion process, especially if the institution had no prior experience in module development and delivery of content in a distance learning environment. Before the pandemic, there was an established trend for online degrees, but we forecast the trend will gain significant attention from young learners as well such as school leavers and career-changing seekers.

Accreditation and degree certification can be used to assure the quality of distance learning degree programs. Given those challenges, the teaching practices need to be published and discussed by the research community. This reflective paper describes one of the popular distance learning degrees from the University of Liverpool (UoL) during the last ten years, as experienced by several of the authors. The paper discusses reflections on many dimensions of the online course, such as class organisation, the evolution of content material, curriculum development, and student satisfaction. It concludes with recommendations for teaching computer security courses for distance learning students, which could also benefit the wider distance learning community.

2. The Reflective Analysis Research Methodology

Reflective practices of teachers and academic staff have been around for some time now, and are a significant activity within the learning cycle (Fathelrahman, 2019). The utilisation of such an activity in
distance learning has proven its importance, as reported by many researchers (e.g. Liu, 2019; Gasparic and Pecar, 2016; Smith and Greene, 2013; Chen et al., 2009; Salmon, 2002). Reflection is considered a crucial mental activity that enables academics to engage more in a meaningful relationship with the academic practices they employ (Mortari, 2015). The hermeneutic perspective of Van Manen (1977) is followed in this paper to cover technical, and practical reflections to improve the quality of distance learning modules, particularly the cyber-security ones. The reflection methodology applied in this paper uses four types of data. Those are the students’ feedback (i.e. responses to the end of module satisfaction survey), the self-generated feedback from the authors while teaching the cybersecurity-related modules, the feedback from course moderators, and the feedback from the program director. As highlighted by Karsten et al. (2019), “For [the] formative purpose, such as improvement of courses over time, it is generally considered best practice to follow this [reflective] method instead of using quantitative tools, such as students’ questionnaires, which are better at assessing individual teacher performance than improving course quality over time”.

3. The Distance Learning Model: Evolution and Illustration

Over the years, our program has evolved several times, while keeping some of its fundamental foundations such as the teaching week structure and the learning management system (LMS). The LMS used is Blackboard (BB), and each module is structured in eight and ten weeks cycles. For an intake module, it is ten weeks, and for a regular module, it is eight weeks. The different components of our online module are depicted in Figure 1, which will be demonstrated throughout this paper.

Until 2014, each week’s resources, including the reading materials, were provided in a textual format, since the majority of the students were from countries where reliable Internet access was not the case. To overcome this hurdle, the decision was made to develop lightweight (i.e. textual) resources that the students could download easily and study offline. The use of multimedia was restricted to links to online YouTube tutorials or similar services. While the idea was appealing back then, the resources were less engaging and less interactive, judging by some of the students’ feedback. In 2014, a decision was made to shift into more multimedia-rich resources, which was in line with the mature stream of using multimedia-rich content at the time (Chen et al., 2009). This includes introductory videos for each week, hand-on tutorials, etc.

![Figure 1. The module Structure and Feedback Loop](image)

It is worth noting that a week-cycle starts on Thursday every week and ends on the following Wednesday, where there are mainly three assessment items. Firstly, the initial Discussion Question (DQ) answer, which is primarily a discussion question that is released before the beginning of the week where the students have to research a topic and compile an answer for. The submission deadline of the initial DQ answer is on Sunday midnight. This simply means the students would have four days of studying/researching the topic and going through the provided resources, before compiling their initial DQ answers. Secondly, the Follow-up posts which the students have to participate in an online forum related to the initial DQ answer where they post 3-5 ‘substantial’ posts in each DQ. Substantial is defined as 1) a critical review of another student’s work and additional information/synthesis that adds or increases the discussion, 2) analysis that includes synthesis of prior week’s work, or multiple
examples in the current week, 3) outside readings that are summarised, 4) reading and comment on another student’s outside readings, and 5) a personal experience/example that applies to the content. Those 3-5 substantial replies need to be submitted on at least three different days out of the seven days of the week cycle. This is the main attendance requirement of the module. It is worth noting that the day of the initial DQ answer submission is not counted if there are no replies posted on that day! Thirdly, the Hand in Assignment (HIA). There is another assessment item which replaces the HIA at the last week (i.e. or weeks) of the module, which is the end of the module project (i.e. sometimes called group project). It is worth noting that for an eight-week module, there are at most 18 assessment points that are distributed evenly, based on the complexity, not the number of submissions as one can see in the sample assessment grid for the Cyber Crime Prevention and Protection module in Table 1.

Table 1. CKIT-530: ASSESSMENT GRID

<table>
<thead>
<tr>
<th>Week</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial DQs and Follow-ups</td>
<td>X 5%</td>
<td>X 5%</td>
<td>X 5%</td>
<td>X(2) 10%</td>
<td>X 5%</td>
<td>X(2) 10%</td>
<td>X 10%</td>
<td>X 5%</td>
<td>50%</td>
</tr>
<tr>
<td>Individual Assignments</td>
<td>X 5%</td>
<td>X 10%</td>
<td>X 5%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X 10%</td>
<td>30%</td>
</tr>
<tr>
<td>Group Project</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>⇒*</td>
<td>X 10%</td>
<td>⇒*</td>
<td>X 10%</td>
<td>-</td>
<td>20%</td>
</tr>
<tr>
<td>Number of Assessments</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>16 %100</td>
</tr>
</tbody>
</table>

⇒*: starts on the current week and is due the next one.

The students provide feedback using a link to a Student Satisfaction Survey after the end of the module, which is instrumental in module evaluation and module continuous improvement (Tricker et al., 2001). Generally, the qualitative research methodology (i.e. using online surveys) is employed in analysing the student satisfaction for distance learning (Kutluk & Gulmez, 2012; Suzek & Ozcan, 2017). In our modules, students would voluntarily and anonymously rate the overall satisfaction with the module content and structure, the interaction with other fellow students and how such interaction made a helpful contribution to the student’s understanding of the module subject, the learning outcomes for the module and whether they are communicated well (Aslam & Rao, 2017), the learning outcomes for the module were met, the consistency of the module structure with other modules, the overall satisfaction with the instructor, etc.

Since this is a reflection paper, which has not originally been designed as a research project, no ethical approval is needed (Karsten et al., 2019). None of the students’ feedback is quoted in this paper; however, as highlighted earlier, the students’ feedback has affected the authors’ reflections. Each instructor is required to respond to a Faculty Feedback request that requires responses on the overall class performance, the occurrences of academic integrity incidents compared to previously taught classes and the steps taken for those incidents to ensure current student evaluation methodology is efficient against plagiarism as investigated in (Çalışkan, 2019). As depicted in Figure 1, after the end of the module, a faculty member carries out the moderation process, which includes assessing the delivery of learning objectives, the application of policies and procedures.

4. Reflective Analysis: Method, Discussions and Recommendations

Designing a module based on a week-cycle (i.e. Waterfall-like model) should assure the learning outcomes are achieved cumulatively and systematically. Some modules, in other educational institutions, only specify the start and the end time of the whole module while offering the full module content all at once. However, in this case, there is no way to make sure that the students progressed seamlessly from the beginning of the module to the end in a sequential manner. Rather, students can start wherever they like. While this provides flexibility, it does not assure the build-up of knowledge in a proper way. For example, a student may find the topic of access control interesting thus jumping to that first, before studying the basics of cryptography, which are extensively used in access control. This creates knowledge loopholes that could be easily sorted out by organising the module in week-cycles. Our observation here is that commitment and motivation are important factors in the success of learning.
online, and it is measured by several factors, including the amount of time a student is willing to devote weekly and the willingness to accommodate the study schedule in their day-to-day activities. Given the recommendation of organising modules in week cycles, two issues need to be considered:

1) It is important to mind the overlapping tasks (i.e. the submission of work while waiting on the reception of the feedback of the previous submission). Module designers should take into consideration quite carefully the time it would take the teaching staff to compile personalized feedback and the time it would take the students to go through and understand that feedback, especially the improvement tips. Failing to observe such a factor (i.e. time) would render the feedback useless.

2) A period of 12 hours or so needs to be considered for those students in the Far East. For example, releasing the discussion questions needs to be 12 hours before the actual due date to accommodate those students 12 hours or so ahead of the time. Besides, marking the student participation and submission should take into consideration the same time window (i.e. 12 hours or so) so that it starts 12 hours or so after the deadline.

Providing constructive feedback for an online class can be challenging for many reasons, such as the diverse backgrounds of the students and the lack of a physical medium to elaborate upon the feedback. Our observation regarding the feedback (supported by responses from the student satisfaction surveys, the module coordinator, and the program director) reveals that a personalised message does not only show what went well and or not during the week but also motivates the students to engage more with the content and contribute actively to discussions in the class.

As the module runs on a weekly cycle, as seen in Figure 2, timely feedback is important. Our experience is that, by policy, the feedback is due no later than three days after the end of the week (i.e. the week cycle ends on Wednesday) thus by Saturday midnight students need to get their previous week’s feedback. This is a bit problematic as the new week starts on Thursday and students need to submit their initial DQ answer before or by Sunday midnight. The preparation takes time, and the expectation is that students may not have enough time to go through the improvement tips in the previous week feedback well before they start compiling their initial DQ answer for the new week. This simply means, students may not accommodate the comments of their instructor in the week following it; at least for the initial DQ answer. Giving the instructor less time to evaluate students’ work as a solution to this problem will affect the quality of the feedback, especially in large classes. The authors attempted to sort this issue out by saving partial feedback in BB rubrics section on the go during the week to not to wait until the end of the week. Unfortunately, the authors realised that BB does not store part of the feedback and all comments saved during the week were gone! Until BB solves this issue or a third-party software does, there is no way but to save those comments off the system (i.e. BB) then consolidate them all at once after the end of the week cycle. To help to mitigate this (i.e. short time for sending the feedback to the students), the teaching teams used a peer assessment mechanism. Ahmed et al. (2020) highlighted that “Peer assessment is considered as a common strategy for evaluating open assignments, increasing learners’ engagement with the educational content and/or for breaking the social isolation some learners might feel during their learning journey”. The details of an innovative method of incorporating peer assessment in the online modules taught by the authors of this paper have been discussed comprehensively by Ahmed et al. (2020) and are outside the scope of this paper.

![Figure 2. The Week Life Cycle](image-url)

The existence of solid rubrics is as equally important as the personalised feedback, and it makes the whole process more transparent for the learners. By the UoL faculty handbook, the assessment
rubrics specify a 6-points grading system (i.e. F: Fail, D: Marginal Fail, C: Satisfactory, B: Merit, A: Distinction, A*: High Distinction. What we observed here regarding the rubrics is while it is comprehensively demonstrating what constitutes a particular grade, the interpretation of those constituents are different from teaching staff to another. The first author as a Core Faculty witnessed cases where various teaching staff record different grades for the same submission. The discrepancy amongst instructors interpreting the marking criteria was considered significant that needed the program director to interject a couple of times to clarify the requirement. This was an unnecessary overhead that, sometimes, delayed the release of grades and, hence, impacted the performance of the students in the new week, since they have not received the instructor comments for enhancement for the previous week. Having said that about the assessment criteria, a decision was made to embed the rubrics inside BB’s Grade Book. Our observation is that embedding the rubrics in the BB grade-book made it easier for the instructors to follow the book and not to get confused about the interpretation of the grading system as provided in the Faculty Handbook. The only concern here is that if an instructor is following the rubrics section and ticks those corresponding boxes without accompanying it with a personalised feedback message, the assessment process becomes mechanical that lacks personalisation of the feedback.

Another important feedback component is the student satisfaction survey (i.e. student feedback on the course), which helps to improve the online modules. The quality of the feedback starts earlier by proposing actions to issues identified in the student feedback and thus closing the loop. Without a systematic process in place to accommodate the issues raised by the students, student feedback loses its significance. The process of accommodating student feedback in module improvement at our institution is somehow blurred. While the Faculty Feedback form requires the teaching staff to reflect on their teaching and to comment on their actions taken to address issues raised by the students there is no enforcement by either the course moderator or the program director to implement (some of) changes to the course. More often than not, the instructors decide not to take any action for the issues from student feedback, and this goes unnoticed by the moderation process. For example, out of date materials and course content will not be considered an actionable item, unless the instructor confirms the need for a material update and the existence of a budget for the change. Another observation regarding the student feedback is the university does not have a methodology to identify biased feedback. For example, feedback by a student that grades are not released on time while grades are being released on time will be normally processed by the Data Analytic team without double-checking. As such reflective feedback is important in evaluating the instructor performance, especially where such ‘anomalies’ need to be identified.

Our final two minor observations consider group formations and the use of multimedia in online classes. Our experience in building student groups is that there are no major differences when forming the groups randomly or based on the time zones. We have had no major issues in the last ten years for group members who are geographically dispersed. We are seeing potential gain within groups with various time zones or backgrounds and skills (Karsten et al. (2019)). Different time zones provide continuity of work in a way that a group member could post his/her contribution overnight to another group member, who will be working on it on the evening of the same day due to time difference. Thus, there is always someone who is working on the project. We acknowledge, however, the challenge this may impose to synchronous communication (i.e. online meetings etc.). Evolving from text-based modules where multimedia is restricted, due to Internet access challenges in the last decade, to the multimedia-rich curriculum has proven to be beneficial to the students, as experienced by the teaching teams and as reflected in the students’ satisfaction survey. Since some instructors are not native English speakers, some video tutorials are hard to follow due to thick dialects. This has been raised by the students many times and needs to be taken into consideration when institutions prepare their video tutorials. As a solution to this problem, the video transcript was embedded in the video for those who cannot follow the video (i.e. including those with a diagnosed hearing disability), which partially sorted out the issue.

5. Conclusions

Delivering successful online courses is challenging due to many reasons. Facilitating the learning of cybersecurity courses are proven to be more challenging due to extensive hands-on nature of these courses, requiring specific infrastructure (i.e. tools, processing power, case studies, etc.). This paper
discussed the program design, evolution, and teaching practices at the university. It also summarised those lessons learnt while teaching computer-security-related modules. Topics such as assessment feedback, assessment rubrics, multimedia-rich content, student feedback, and the overall week-based modules are discussed. The quality of assessment feedback and the existence of clearly defined assessment rubrics are major factors in offering successful online learning experiences. Personalised, timed, and positive-toned feedback are amongst other attributes that contribute to the success of teaching online. Our observation in the past ten years is that personalising the feedback enhances its quality and further helps with establishing a positive relationship with the learners and improves social interaction (Wu He, S., 2014). On-time feedback is also important especially for modules structured as week cycles. To assure fairness and reduce grade challenges, the assessment rubrics need to be embedded in the module’s Grade Book. Exemplary cases need to be shared frequently amongst the teaching team to assure consistency across different instructors, teaching different modules.

References

Using Online Learning Environments to Address Digital Literacy Competencies of Construction Management Graduates

Thilini JAYAWICKRAMA *, Shokry ABDELAAL * & Rhodora ABADIA *
*UniSA Online, University of South Australia, Australia
*Thilini.Jayawickrama@unisa.edu.au

Abstract: Digitalization has become popular across various industries including construction. Construction industry involves unique outputs, complex processes and multi-disciplinary engagements. Digitization has introduced in the industry from electronic documents, email communication to automation and concepts such as Building Information Modelling (BIM). Construction professionals should possess sets of skills to cope with complex nature in the industry whilst conquering this modern digital era. United Nations Educational, Scientific and Cultural Organization (UNESCO) has proposed a set of digital literacy competencies for youth and adults in their ‘Digital Literacy Global Framework’. Meanwhile, some professional bodies in construction also have acknowledged Information and Communications Technology (ICT) skills as core to their professionals. In this vein, digital literacy will be a competitive advantage for construction management graduates. Digital literacy practices are not emerging spontaneously but, education plays an important role in this context. Learning Management Systems (LMS) are widespread among higher education providers for both face-to-face and online offerings. LMS are tailored to cater various pedagogical strategies and can also be used to enrich digital literacy of students. This paper maps how the digital literacy competences proposed by UNESCO are addressed with online activities and resources used in first year courses in a construction management degree program operating in an online educational environment in Australia facilitated through Moodle Learning Management System. The core competencies recognized by major construction professional bodies were considered as the ‘Career-related competences’ in the UNESCO framework. The paper also presents which of the online activities are mostly implemented and engaged in the selected courses and which of the core competencies are mostly addressed. Findings suggest ‘Assignment’, ‘Forum’ and ‘Quiz’ as the mostly implemented and engaged activities, whereas ‘Glossary’ activity with less implementation but, higher engagement so, with a higher potential of addressing digital competencies. Information and data literacy, communication and collaboration, digital content creation and career-related competency of using standard computer packages are the mostly addressed competencies.

Keywords: Construction Management, Digital Literacy, Moodle Activities

1. Introduction

Digitalization in construction industry has evolved throughout the years moving beyond electronic document management (EDM) systems to advanced techniques of virtual construction prototyping with Building Information Modelling (BIM) (JR Jupp & Ramsey Awad, 2013). The resultant move from traditional ways to digital ways of working posed tensions on the construction manager’s role demanding new knowledge and skill base (Harty & Whyte, 2010). Ezcan, Goulding, and Arif (2020) identified training and education as one of the drivers for technology diffusion-adoption in the construction industry.

The way construction management students learn have also changed from studying face-to-face, to blended learning, and increasingly fully online. While digital literacy skills facilitate students to engage in e-learning modes (Mohammadyari & Singh, 2015), conversely there are studies [e.g., Meurant, 2010; Martinez-Alcala et al., 2018; Patmanthara & Hidayat, 2018] investigating the role of digital learning modes in enhancing digital literacy skills of learners. There have studies that looked
at different approaches in developing digital literacy in construction management. Some studies looked at how design thinking can be used in developing digital literacy in construction management and looked at the use of BIM (Forsythe, Jupp, & Sawhney, 2013; Julie Jupp & Ramsey Awad, 2013). Given the increasing number of students studying construction management degrees online in construction management, there is a need to identify how online teaching and learning activities can address digital literacy competencies required by construction management students.

Computer-based, internet-hosted Learning Management Systems (LMS) provide online delivery platforms for educators and learners to share learning resources and engaging with useful techniques to foster active learning (Tang & Chaw, 2016). The free open-source Moodle is one of the fast growing LMS offering a range of tools for resource sharing, discussions, calendar planning, assignment submissions and grading (Deng & Tavares, 2013). The design and development of learning activities can be done with the use of these various functions effectively by considering skills and knowledge base required by learners. The standard Moodle offers seven (07) types of resources that teachers can use to support learning by adding content to their courses. In contrast, activities enable students to contribute directly, interact with other students or the teaching team and Moodle offers 14 types of such activities.

This paper aims to find the most relevant Moodle techniques to address digital literacy competencies required by the construction management undergraduate students. The paper maps online activities and resources offered by Moodle with competencies proposed by UNESCO including career-related ICT competencies recognized by major construction professional bodies.

2. Digital literacy competencies for construction managers

The United Nations Educational, Scientific and Cultural Organization (UNESCO) proposed six competence areas with digital literacy competencies under each area in their ‘Digital Literacy Global Framework’ (DLGF) serving the Sustainable Development Goal (SDG) thematic Indicator 4.4.2: “Percentage of youth/adults who have achieved at least a minimum level of proficiency in digital literacy skills” (Law, Woo, and Wong, 2018). These competency areas are namely; devices and software operations, information and data literacy, communication and collaboration, digital content creation, safety, problem-solving and career-related competences. For construction-specific careers, some major professional bodies in construction have specified core ICT skills and competencies for their professionals.

The Australian Institute of Quantity Surveyors (AIQS) specifies three competencies under ‘computer services’ as a core unit under ‘support competencies’ in their competency standards document for quantity surveyors, construction economists and cost engineers. The first competency serves general usage of computers including general skills in the use and care of computing hardware and software, using computers to access, enter and process information, and data between participants in construction activities. These are addressed in the UNESCO framework under competency areas of ‘devices and software operations’, ‘information and data literacy’, and ‘communication and collaboration’. The second competency specified by AIQS addresses the usage of standard application packages such as word processing, electronic spreadsheets and statistical packages in the management and presentation of information related to construction activities. The third AIQS competency serves the usage of construction-specific computer packages. Second and third AIQS competency areas will be considered under career-related competences in this study.

‘Requirements and Competencies guide’ document published by Royal Institution of Chartered Surveyors (RICS) specifies ‘data management’ as a mandatory competency including; obtaining and using published sources of data, managing inhouse sources of data and the use of computerised central project databases or BIM. This will be considered under career-related competences in this study. Table 1 presents the detailed descriptions of the competencies considered in this study.

<table>
<thead>
<tr>
<th>Competency 1 - Devices and software operations</th>
</tr>
</thead>
<tbody>
<tr>
<td>To identify and use hardware tools and technologies.</td>
</tr>
<tr>
<td>To identify data, information and digital content needed to operate software tools and technologies</td>
</tr>
</tbody>
</table>
Competency 2 - Information and data literacy
- To articulate information needs, to locate and retrieve digital data, information and content.
- To judge the relevance of the source and its content.
- To store, manage and organise digital data, information and content.

Competency 3 - Communication and collaboration
- To interact, communicate and collaborate through digital technologies while being aware of cultural and generational diversity.
- To participate in society through public and private digital services and participatory citizenship.
- To manage one’s digital identity and reputation.

Competency 4 - Digital content creation
- To create and edit digital content.
- To improve and integrate information and content into an existing body of knowledge while understanding how copyright and licenses are to be applied.
- To know how to give understandable instructions for a computer system.

Competency 5 - Safety
- To protect devices, content, personal data and privacy in digital environments.
- To protect physical and psychological health, and to be aware of digital technologies for social well-being and social inclusion.
- To be aware of the environmental impact of digital technologies and their use.

Competency 6 - Problem-solving
- To identify needs and problems and to resolve conceptual problems and problem situations in digital environments.
- To use digital tools to innovate processes and products.
- To keep up to date with the digital evolution.

Competency 7 - Career-related competences
- Use standard application packages in the management and presentation of information relating to discipline specific activities
  - Word processing programs to produce professional quality reports,
  - Electronic spreadsheets used to prepare schedules
  - Statistical packages use to manage and process statistical data
- Use computer packages for construction-specific disciplines (e.g., CostX, Revit)
- Data management
  - Obtain and use published sources of data
  - Collect, analyze, store inhouse sources of data
  - Use of computerised central project databases or Building Information Modelling (BIM)

3. Methodology

3.1 Data

The online construction management degree is one of the most popular courses in University of South Australia (UniSA). UniSA is known as one of the largest providers of online learning in Australia. As part of its digital strategy, UniSA Online was created to offer a variety of degrees from different disciplines that are 100% online. For the construction management degree, around 70% of the students studying are working in the construction industry and would want to progress their career or shift their career in the field of construction management.

The data in this study was collected from seven first year courses in construction management degree program delivered from January 2019 to March 2020. These courses can be categorised under major steams of construction management, construction technology, construction cost management and construction law. These courses were developed with Moodle resources such as files, pages and URL. The Moodle activities implemented in these courses are forums, assignment, quiz, database, survey, glossary, lesson and Zoom. Although Zoom is not a standard Moodle activity, this is embedded in the Moodle and used in the online construction management courses. Since the online courses are designed to be asynchronous, Zoom sessions are recorded for students to view if they missed the live session or
for them to watch again for revisions. These activities and resources enable to generate reports of student engagement.

This study primarily used the student engagement reports which provide the percentage engagement in each activity and resource in a course based on views and/or contributions. The engagement activity of 781 online construction management students were collected. Each selected course was reviewed for the usage of various Moodle activities and resources in a focus group discussion with individual Online Course Coordinators. A total of 1248 records in the student engagement report were collected. Techniques dedicated to the online teaching team were omitted to filter only relevant activities and resources. Hence, out of 1248, 345 records of activities were removed. These include activities and resources that were hidden to the students (e.g., activities for the online academic staff, alternative activities that were not used) but still recorded in the student engagement report.

3.2 Focus Group

Online Course Coordinators from the seven online construction management courses participated in the focus group. The purpose was to identify how each of the activity and resources are used in the courses in the context of construction management. Feedback from the Online Course Coordinators were noted. Open coding was used to take note of margin of words, theories or short phrases to sum up what is being said. The words and phrases from the Online Course Coordinator’s responses were collected and duplications were removed to eliminate overlapping of categories on how they use the activities and resources.

3.3 Data Analysis

The list of online activities and resources, taking into consideration the result of the focus group interviews, were mapped against digital literacy competencies defined by UNESCO and the construction professional bodies (AIQS and RICS). The student engagement reports were used to analyse the extent each activity and resource implemented and engaged in courses. Findings suggest which of the Moodle activities and resources were highly relevant in addressing digital literacy competencies of construction management students.

4. Results and Discussions

4.1 Online Activities and Resources implemented in selected courses

This section discusses implementation of Moodle activities and resources in the selected courses as identified by the Online Course Coordinators during the focus group discussion. Types of courses implemented each activity and resources with the intended purpose are presented in Table 2. In the study, the data for files, pages and URL are aggregated because these activities are used for the same purpose of providing information to the students.

<table>
<thead>
<tr>
<th>Online Activity/Resource</th>
<th>Course/s</th>
<th>Purpose</th>
</tr>
</thead>
</table>
| Course website (Moodle)  | All courses | • Course website is created using Moodle  
• Minimum computer requirements are specified by the university including; Strong internet connection, hardware (Webcam, Microphone, Headset, Speakers) |
<p>| Moodle Files, Pages and URLs | All seven courses | • Used to share content with students |
| Moodle Forums            | All seven courses | • Used as the major communication mode within course website |</p>
<table>
<thead>
<tr>
<th>Activity</th>
<th>Courses</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moodle Assignment</td>
<td>All seven courses</td>
<td>• Allow students to have asynchronous discussions with peers or academic team</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Used as a summative activity for students to submit assignments online (this activity is used as the major submission portal for most of the assignments)</td>
</tr>
<tr>
<td>construction cost management</td>
<td></td>
<td>• Used as a formative activity for students to submit progressive hurdle activities and receive model answers</td>
</tr>
<tr>
<td>Moodle Quiz</td>
<td>All seven courses</td>
<td>• Used as a formative activity for students to test their understanding of a concept and get feedback automatically</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Used as a summative activity to assess competencies of students at a certain stage of a study period</td>
</tr>
<tr>
<td>Moodle Database</td>
<td>construction cost management</td>
<td>• Used as a formative activity to upload final output after following a series of tasks, view and comment on others works</td>
</tr>
<tr>
<td></td>
<td>courses</td>
<td></td>
</tr>
<tr>
<td>Moodle Survey</td>
<td>Two courses</td>
<td>• Used as a formative activity to survey the early preparedness of students with a software from an external provider</td>
</tr>
<tr>
<td></td>
<td>construction cost management</td>
<td></td>
</tr>
<tr>
<td></td>
<td>courses</td>
<td></td>
</tr>
<tr>
<td>Moodle Glossary</td>
<td>construction management</td>
<td>• Used as a formative activity to contribute to a shared glossary defining key words found in weekly concepts</td>
</tr>
<tr>
<td></td>
<td>courses</td>
<td></td>
</tr>
<tr>
<td>Moodle Lesson</td>
<td>One course in construction cost</td>
<td>• Used as a formative activity to follow a series of tasks following examples and receiving answers automatically</td>
</tr>
<tr>
<td></td>
<td>management</td>
<td></td>
</tr>
<tr>
<td>Live sessions via Zoom</td>
<td>All seven courses</td>
<td>• Facilitate interaction between the online academic team and the students</td>
</tr>
</tbody>
</table>

It was found that some activities are commonly implemented in courses but, some are implemented occasionally. Only the commonly implemented activities and resources namely, files, pages, URLs, forums, assignment, quiz, database, glossary and Zoom were considered for mapping with competencies and further investigation with student engagement data.

### 4.2 Mapping of Competencies

This section maps the contributions of selected Moodle activities and resources (Table 2) in addressing digital literacy competencies specified by UNESCO and career related competencies specified by AIQS and RICS. This was performed through the Online Course Coordinator’s focus group interviews and summary of results is presented in Table 3. This mapping considered which activities and resources can address each competency based on the detailed descriptions of each competency presented in Table 1.

It is evident that competency 2 is supported by all Moodle activities and resources implemented in the courses. Competency 3, competency 4 and the career-related competency on using standard computer packages are also mostly supported by Moodle techniques. Other competencies are supported with only two among different techniques. Online Course Coordinators’ views on the contributions of each activity and resource towards identified competencies are discussed hereafter.
Table 3: Contribution of Moodle Activities and Resources Implemented in Selected Courses.

<table>
<thead>
<tr>
<th>Digital Literacy Competencies</th>
<th>Moodle course site</th>
<th>Forums</th>
<th>Quizzes</th>
<th>Files, Pages, URL</th>
<th>Database</th>
<th>Glossary</th>
<th>Assignment</th>
<th>Live online sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Competency 1 - Devices and software</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Competency 2 - Information and data literacy</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Competency 3 - Communication and collaboration</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Competency 4 - Digital content creation</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Competency 5 - Safety</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Competency 6 - Problem-solving</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Competency 7 - Career-related competences</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

4.2.1 Files, page, URL

Apart from intext body of the course website, files, pages and URL are the mostly implemented Moodle resources to share course content in an organized manner in all courses. Students watch or download these contents from the course website and save for studying offline where necessary. Also, they download and save assignment resources which are mostly shared as files to work offline. These activities directly contribute to their information and data literacy (competency 2). Moreover, in this construction management program, most of the course resources are shared as Word documents, Excel spreadsheets and other files produced by construction-specific computer packages (e.g., CostX files). This has a direct contribution to career-related competencies where construction management students are expected to be competent in using standard application packages such as Word and Excel, and construction-specific computer packages. Moreover, RICS specified data management as a core competency and students in this program practice to access and use published sources of data via URLs provided in most courses for relevant contents.

4.2.2 Forums

As the major mode of communication with peers and academic team, students post and upload digital content (files, images) in forums, read posts to find relevant information and sometimes respond to the posts by peers or academic team. Meurant (2010) also explained forum activity as generating animated discussions and therefore, develop students’ ability to navigate, engage, contribute in sustained written discussions and debate. Therefore, forums directly contribute to competency 3, communication and collaboration. At the same time, forums contribute to improve their information and data literacy (competency 2) through navigating and engaging to create posts, read and respond. Moreover, students can activate Moodle option to receive email notifications for each forum post or a daily digest for selected forums for receiving up-to-date information effectively thereby, judge the relevance of the source and its content as given under competency 2. Another popular use is dedicated forums for communicating the technical issues on accessing course contents or other course-related applications to seek solutions and to inform lessons learned. These applications include construction-specific computer packages such as CostX and Revit. Some issues are resolved collectively with peers and teaching teams thereby, contributing to improve the problem-solving skills under competency 6.
4.2.3 Assignment

As mentioned in Table 2, this Moodle activity is implemented in all courses as the major submission portal for most of the assignments requiring students to produce digital content offline in a specified file format (documents, files using other applications, photo images, videos or presentations) and upload to submit. Thereby, this activity has a direct contribution to improve their information and data literacy competencies (competency 2) and digital content creation competencies (competency 4).

These answer files can be in different file formats and similar to the findings discussed under files, pages and URL resources, construction management students are producing these answers using standard application packages and construction-specific computer packages. Thus, having a direct contribution to career-related competencies of construction management students. Moreover, students are monitored, and actions will be taken against academic integrity issues based on the Turnitin report created in Moodle assignment portal for these answer files. This process will inform and enhance understanding of students on how copyright and licenses are to be applied as also required under competency 4.

4.2.4 Quizzes

As can be seen in Table 2, quizzes are used as both summative and formative activities in the courses and as presented in Table 3, Online Course Coordinators pointed out a range of skills developed through quizzes by navigating over a quiz to read, understand, answer, flag and review various types of questions within a given duration. These types of questions require students to select answers using automated options; type answers online; and/or download resources and upload completed answer files. Thereby, quiz activities can be used to address information and data literacy (competency 2) of students. Moreover, students need to give various commands in the online quiz environment to answer above types of questions, flag, review and to use submission options. Therefore, students will develop competency 4 up to a considerable extent by practicing, creating and editing digital content, and by giving understandable instructions to the system. The resources and answer files can be in different file formats and similar to the findings discussed under files, pages and URL resources, construction management students are mostly required to use Word documents and Excel spreadsheets in these quizzes as required for essay type questions.

4.2.5 Database

This activity is implemented in few courses where students need to navigate over a series of tasks to produce a digital output (documents or photo images), upload, review other works and respond. Therefore, this activity can contribute to information and data literacy competency of students (competency 2). Also, it contributes to communication and collaboration (competency 3) due to the feature allowing students to review and respond to others works. While producing those digital output, students will be directly benefited by addressing competency 4. As similar to other activities in this program, most of these digital outputs contain Word and Excel files thereby, practice using standard applications as required to address career-related competencies. However, all the Online Course Coordinators found it is not user-friendly given the tasks in construction management courses.

4.2.6 Glossary

This activity is not implemented in many courses but, Online Course Coordinators agreed that it can contribute to competency 2 and competency 4 as students need to search definitions of key words, create own descriptions to share in a collaborative space. Given this collaborative environment it can directly contribute to communication and collaboration (competency 3) with contributions from all students. Moreover, Online Course Coordinators suggested that this activity can contribute to data management competencies in a collaborative environment similar to construction projects and thus, will contribute to career-related competencies.
4.2.7 Live sessions using Zoom embedded in Moodle course website

As the major active communication mode, students need to access Zoom application for all group and individual meetings and discussion sessions. During the sessions, they need to manage Zoom application as per their choice either to listen passively (video off, mute) or actively engage using required hardware (speakers, webcam, mic). Thus, it has direct contributions for students to improve their competencies to identify and use software and hardware applications (competency 1). Students who are not able to attend the live session or those doing revision with session, watch the recordings and seeking further information relevant to major course content and thereby, addressing information and data literacy up to some extent (competency 2).

These sessions are indeed the major mode to address communication and collaboration (competency 3) among online students where they interact in a live digital environment with peers and academic team through formal or informal discussions. At the same time, Zoom sessions are created within course website for one-off or recurring sessions throughout a study period, and use with password protection. In this process, students are informed the requirement of protecting privacy in digital environment and thus, contribute to the competencies under safety (competency 5). The academic team use Zoom sessions to assist students seeking solutions for technical issues on course-related applications via ‘screen share’ and ‘remote control’. Some issues are resolved collectively in the presence of peers. This creates problem-solving situations in digital environments and contributing to the competency 6.

4.3 Implementation of Moodle Techniques and Student Engagement

This paper aimed to find the most relevant Moodle techniques to be implemented to address digital literacy of construction management students. Effectiveness of these techniques is not able to achieve through mere implementation of activities and resources in the courses but, student engagement is also important. Figure 1 presents average implementation of each technique in the selected courses and overall student engagement.

![Figure 1: Implementation of Moodle Techniques and Student Engagement](image_url)

Results show file/page/URL resources as the mostly implemented technique in courses. This can be attributed with the fact that most course materials are shared in these forms apart from intext content and concept videos. Student engagement is also considerable as some of these resources provide useful course content and assessment details.
Among different Moodle activities, forum is the mostly implemented activity and showing a considerable student engagement. Findings reported by Deng and Tavares (2013) suggest that social relationships and owning by peers as the motivational factors for students to engage in online discussions. Their study indicates the requirement of designing and facilitating forums to address the needs of students from informal chatting to discussion of academic issues. This can be considered by the academic team in the construction management courses to further improve student engagement because, communication is important for construction management students whereby construction industry is multi-disciplinary in nature and require vast communication and negotiation among stakeholders.

Despite the less implementation, assignment activity has the highest student engagement given it is the major mode of submitting summative assessments. This activity contributes to several digital competencies and involve outputs produced by a variety of computer packages thus, an effective Moodle activity for construction management courses.

Quizzes are fairly implemented in courses for both summative and formative activities, and summative activities are compulsory for students. However, all Online Course Coordinators agreed that quizzes are popular among students even if used as formative activities as reflected in Figure 1. Meurant (2010) also identified that if the quizzes are set accordingly, the student can find out their score immediately on completion which is a significant advantage. This can be the reason for high engagement in quizzes despite the activities are being summative or formative. More quizzes with automated marking can be implemented in courses as an effective technique with popularity among students and involve a range of skills with a variety of question types and thus, contributing to several digital literacy competencies.

Databases activity has fair implementation but, lack of student engagement reflecting the view of Online Course Coordinators regarding lack of user-friendliness. Glossary activity has lack of implementation but, higher student engagement and more contributions towards digital literacy competencies. Therefore, this study suggests implementing Moodle glossary in more activities to gain its benefits but, to limit activities with databases or to explore more effective tasks to implement with.

Results show lack of engagement in live sessions (participation) but, comparatively higher engagement in viewing recorded sessions. This is somehow in contrast with the view of Abdous and Yen (2010) stating live class meetings as an increasingly important delivery method in online learning given the less effectiveness among these online students. Anyhow, the effectiveness of live sessions cannot be undermined as this an online student group studying part time and engage in asynchronous learning from interstate. Personal commitments and different time zones attribute to the less active engagement but, more views of the recording. It is worth to explore strategies to increase the active participation with this activity as it contributes to several competencies including the use of software and hardware and, more importantly contributing to improve communication skills of construction management students learning online.

5. Conclusion

This paper reviewed the ‘Digital Literacy Global Framework’ proposed by UNESCO and construction professional bodies (AIQS, RICS) to identify the digital literacy competencies required by construction management graduates. It also reviewed online activities and resources offered by Moodle Learning Management Systems with the implementation in seven first year courses in a construction management degree program operating in an online educational environment in Australia. Identified digital literacy competencies were mapped with Moodle techniques implemented in selected courses as a result of a focus group discussion with Online Course Coordinators. This mapping identified that information and data literacy competency is addressed with all implemented Moodle techniques. Findings suggest ‘Assignment’, ‘Forum’ and ‘Quiz’ as relevant and therefore, recommended to implement for the potential enrichment of digital literacy of students. ‘Glossary’ activity found with less implementation but, as a relevant technique to be used due to higher engagement and higher contributions to general and career-related competencies. Supported with literature findings, the study suggests activities such as forums and live sessions should be implemented carefully to retain student engagement as these are very important for construction managers given the nature of industry with communication and collaboration with multi-disciplinary stakeholders. The current study recommends the relevant online activities and resources with high relevance to address digital literacy competencies.
based on the level of implementation and student engagement and this study can be extended to find the effectiveness in enriching digital literacy of construction management students in the online learning environment.

References


E-book based Learning in times of Pandemic

Rwitajit MAJUMDAR *, Mei-Rong Alice CHEN *, Brendan FLANAGAN * & Hiroaki OGATA *

* Academic Center for Computing and Media Studies, Kyoto University, Japan
*dr.rwito@gmail.com

Abstract: As a necessary preventive measure against the spread of Novel Coronavirus (COVID-19), many educational institutions across the globe are changing the style of traditional teaching-learning interactions from classrooms to online modes in a very short period. Data gathered from an online survey conducted with participants (educators and researchers, n=57) from 19 different countries, found that digital divide, uncertainty regarding student's participation in the online activities, and the increase in effort of the teachers to make the rapid transition to a new online mode of teaching and learning are some of the common challenges during this period. We propose how BookRoll, an eBook-based and learning analytics enhanced system has a low floor and high ceiling to continue interactive learning activities during this period. This proposal has the following advantages: It takes lower bandwidth to distribute learning materials than video-based learning activities; teachers can directly use their previously created content with the students and provide easily trackable in-content formative evaluations (specific reflective questions or annotation-based activities).

Keywords: Learning Analytics, eBook-based teaching and learning, Learning Evidence Analytics Framework (LEAF), BookRoll, K-16 education, COVID-19 pandemic

1. Learning Status during Pandemic and Motivation

While COVID-19 continues spreading, like in many countries, Japan too enforced emergency remote teaching and learning at schools and universities. In this work we explore the affordances of eBook technologies to support this challenging period. A questionnaire-based survey was created to collect the perception of educators regarding the current situation and issues in migrating to online pedagogy to continue education. It was distributed by purposive sampling, through a mailing list and social media post from 17 March 2020. The mailing list had 513 members involved in learning analytics research and practice at 272 institutions in over 20 countries. The members of this community were selected as they would be aware of digital teaching-learning scenarios in their country, and some are involved in policy making at institutions. 57 responses were received by 6 April, 2020 from 19 different countries. A majority, 48 (84%), identified as university teachers, and 32 (56%) were associated with higher education focused educational technology research. 8 were in administrative roles at universities and 3 from K12 schools. 32 (58%) reported they had both synchronous and asynchronous online components for teaching-learning activities. Teachers often create or curate their teaching materials (notes, assignments, assessments) and conduct these online activities. Amongst the respondents, 51 (89%) created their lecture materials using presentation software and added voice, 4 of them while conducting synchronous online lecturing. 21 (36%) recorded their lecture for delivery using online software and also using screen recording. For synchronous teaching 3 of them used online whiteboard and polling apps. 17 (30%) of the teachers were preparing to let students view learning materials, participate in discussion forums, submit assignments and take assessments online. The degree of support given by institutions to the respective teachers included guidelines, resources to create materials and distribute among learners, staff support for creating the materials and training sessions before or during this period. While 12 (23%) respondents reported that they had all support items, 10 (19%) only received guidelines from their respective institutes. As a teacher they expected their students were using home’s broadband internet connection accessing via PCs or Tablet. Mobile phones used mobile data. Many expected the duration of online activities of the students to be more than 5 hours per day.
2. **BookRoll: A Learning Analytics enhanced eBook platform.**

BookRoll can be linked to most learning management systems (LMS) by LTI and teachers can upload learning materials in PDF format, and students can access them in a wide range of devices through a standard web browser. The platform then records the interactions of browsing and annotations on those materials to collect log data. Figure 5 shows the reader’s interface and functions in BookRoll. Users can annotate by marking sections of reading materials in yellow to indicate sections that were not understood or red for essential sections. Students can type or draw memos on a page or with a marker to attach it to a specific section of the page. Users can also bookmark pages or use the full-text search function to find the information they are looking for later when revising. Based on the finding in the survey, we also developed an audio upload function associated with each learning material (see the interface in Figure 6). Now teachers can easily upload their spoken lectures or any other tutorial audio associated with those materials. The students can control the audio, which plays automatically when the material is accessed. Based on analysis of the log data LEAF provides various services for the users, such as a learning dashboard for monitoring reading behaviors (Majumdar et al. 2019) to conducting specific teaching-learning activities, AI-driven content recommendation systems (Yang et al. 2019), and reading supports such as smart dictionaries (Lecailliez et al. 2020).

![Figure 5. The reader’s interface in BookRoll system](image)

2.1 **BookRoll features and teaching-learning strategy to resolve some of the challenges.**

**It is a challenging situation - difficult for many but accepted by all:** While the thrust of digitized learning materials makes it a core part of modern formal education here in Japan, the migration was gradual. Nevertheless, during this time, many teachers accepted and created their materials and used available resources for their course using BookRoll. In K University instructors for more than 140 courses signed up to use BookRoll for their course in the Spring 2020 semester. At school, the system usage went up to more than 1000 unique users logging in every day after 7 May.

**Connectivity and Digital divide issues:** Connectivity and the digital divide is an issue even in Japan. While the infrastructure was available in the Japanese school and universities, the same may not be the case at homes of the teachers and students. For schools, they issued nearly 360 Microsoft Surface tablets to all Junior High school students on the only day that the school opened in April. While this would provide the device, still the students required to have a stable internet connection at home to access the learning management system and BookRoll linked to it. A team supported the creation of accounts for all the students and verified network accessibility to the learning tools and materials. To tackle unfamiliarity, face to face teacher training was conducted for the school earlier this year. At universities, a series of online synchronized demonstrations informed the functionalities to participating instructors highlighting the new ones of uploading audio clips along with the reading material.

**A strategy to use BookRoll:** Based on the existing functionality, a teaching-learning strategy is proposed that can be adopted quickly while using BookRoll and the learning dashboard LAViEW. An overview of system architecture, along with the teaching-learning strategy, is presented in Figure 7. Teachers can create their material and save them as PDF materials and the lecture or voice commentary as an audio file. When uploaded to the BookRoll platform, it is accessible from the learning management system. The registered materials can be accessed only by the enrolled students of any course, and the course instructor can control whether the students can download any material.
This feature helps the teachers to provide selected access to their copyrighted materials only with their registered students. The students then can read the material and hear the associated sound clip. Teachers can additionally give annotation-based exercises such as Topic-scanning guiding strategy for English as foreign language learning (Chen et al. 2019) or design learner centric approaches (https://lcm-model.org/) for their online class. The student's annotations and reading behaviors can be reviewed on LAViEW, based on which remedial activities can be provided to the students if required.

3. Conclusion

Scholars have started observing this problematic period from the perspective of educators with different labels such as Emergency Remote Teaching (ERT) Hodges et al. (2020) or Panic-gogy (Kamenetz A., 2020). In this article, digital divide, uncertainty regarding student's participation in the online activities, and the increase in effort of the teachers to coordinate as well as modify their practices in a short period to a new online mode of teaching and learning are some of the common challenges during this period as found from the analysis of the response. This survey-based study captured an initial response of the educators to attempt and understand the problems in technology-enhanced learning during the dynamically evolving COVID-19 crisis. However, the findings helped to reflect on the technology affordances of the eBook-based teaching-learning platform during this pandemic period at least for the higher education sector. There remains further scope of investigating the issues faced by the teachers at K-12 level. Our current proposal of using BookRoll for e-books and audio-based learning has two main advantages to assist easy transition to online mode from the traditional face to face classes. It takes lower bandwidth to distribute learning materials than video-based learning activities and teachers can directly use their previously created content with the students. Further it provides easily trackable in-content formative evaluations using reflective questions or annotation-based activities.

Acknowledgements

This research was supported by JSPS KAKENHI 16H06304, 20K20131, 20H01722, SPIRITS 2020 of Kyoto University and NEDO Special Innovation Program on AI and Big Data 18102059-0.

References


Majumdar R., Akçapınar A., Gokhan Akçapınar, Flanagan B., Ogata H., Learning Analytics Dashboard Widgets to Author Teaching-Learning Cases for Evidence-based Education Authors (2019), in the companion proceedings of the 9th LAK, Tempe, USA pp.600-607


Implementation Challenges of STEM Education: 
from Teachers’ Perspective

Xiaojing WENG a*, Morris S. Y. JONGb , Thomas K.F. CHIUb

abcDepartment of Curriculum and Instruction, The Chinese University of Hong Kong, Hong Kong 
*a+weng@link.cuhk.edu.hk

Abstract: How to effectively support teachers in implementing STEM education is regarded as an urging issue in school education. To address this issue, we need to gain a better understanding of teachers’ implementation challenges of STEM education. This qualitative research collected views from the teachers who participated in a STEM education training programme. After the programme, four teachers took part in the semi-structured interviews. The study revealed that their lack of (i) pedagogical content knowledge (PCK) and (ii) external support were the main challenges that they were facing. Suggestions to facilitate teachers’ STEM education implementation in schools, as well as research implications, are proposed and discussed.

Keywords: STEM education, implementation challenges, school education, teachers’ views

1. Introduction

Literature suggested that teachers have faced many challenges and obstacles when implementing STEM education (Geng, Jong, & Chai, 2019; Lesseig et al., 2016; Seals et al., 2017; Shernoff et al., 2017). Teo and Ke (2014) pointed out several challenges of STEM teaching in different Singapore schools. The challenges included creating a new subject from scratch and using new technologies in teaching. Shernoff et al. (2017) interviewed 22 teachers to understand what they need from STEM teachers’ professional development activities. They listed out teachers’ difficulties, including lack of (i) pedagogical content knowledge (PCK) and content knowledge, (ii) time for collaboration and planning, and (iii) knowledge about assessment. Additionally, Geng et al. (2019) have surveyed teachers’ self-efficacy and concerns about the implementation of STEM education in Hong Kong. According to their study, teachers’ challenges included insufficient PCK for interdisciplinary learning and lack of instructional designs promoting engineering discipline. The role of teachers in STEM education is very important as they are the designers. We need to better understand the challenges that teachers are facing for better STEM education implementation (Shernoff et al., 2017). The main goal of this study is to investigate the challenges that STEM teachers have faced in Hong Kong, and in turn, to propose some possible suggestions for addressing the challenges. The core research question of the study is: “What challenges do Hong Kong teachers face when implementing STEM education?”.

2. Research design

2.1 Research participants

Four participants took part in the research, see Table 1. They came from a STEM training programme organized by a research-intensive university in Hong Kong. In the programme, there were three instructors from the university and 49 teachers from the schools. The teachers had a diversified teaching background, including Chinese/English language education, mathematics education, general studies, and visual arts, and shared a common goal to learn more about STEM education. We adopted the convenience sampling method to approach the participants. Interview invitations were sent to all the learners after their STEM training programme.
Table 1. Interviewees’ Information

<table>
<thead>
<tr>
<th></th>
<th>Teacher 1</th>
<th>Teacher 2</th>
<th>Teacher 3</th>
<th>Teacher 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seniority</td>
<td>18 years</td>
<td>14 years</td>
<td>8 years</td>
<td>28 years</td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
<td>Male</td>
<td>Male</td>
<td>Male</td>
</tr>
<tr>
<td>Major</td>
<td>Physical</td>
<td>Business</td>
<td>Psychology</td>
<td>Computer</td>
</tr>
</tbody>
</table>

2.2 Research procedure

This study adopted a qualitative approach to investigate the research question. We used semi-structured interviews to collect the view of participants. Each interview lasted about 45 minutes. We got the participants’ consents to audio-record the interviews. The teachers were interviewed either at the teachers’ schools or the university. The researcher has adopted an interview guide for the interviews. The interview guide, which was developed by two researchers from the university, aimed to collect teachers’ concerns about implementing STEM education in schools. A question like “What challenges do you have when implementing STEM education in schools?” is included in the interview guide. All the interview data were transcribed verbatim and analyzed with the guidance of grounded theory (Lambert, 2019) in the later steps of the research.

3. Results

3.1 Lack of PCK

The teachers found that they did not have sufficient PCK and related teaching experience to implement STEM education. They expressed that they were not well prepared and should improve their related teaching skills and understanding of STEM learning for better implementation. See the excerpts: “I am worried about teachers’ capabilities to implement STEM education practice at the current stage” (teacher 1). “The schools do not have STEM experience, so their teachers may not understand what STEM is” (teacher 3). The teachers expressed that there were insufficient teacher training workshops for them to improve their PCK. “The teacher education institutions in Hong Kong should take their responsibilities to help teachers develop their STEM teaching skills” (teacher 1). “If there is no related professional training for teachers, they can hardly teach STEM. We need more teachers who can teach in the STEM area; it is the first task to implement STEM education” (teacher 3).

3.2 Lack of external support

The teachers perceived that lacking external support was a challenge for them to implement STEM education. The support should allow teachers to exchange teaching ideas, learning materials, STEM resources, as well as to get feedback from teachers in other schools, students, parents, and other stakeholders. In other words, the teachers lacked STEM teaching communities. See the excerpts: “I think teachers are facing many constraints from the external environment that stop them from implementing STEM education” (teacher 2). “I felt the community of practice in STEM education has not been established. Ideally, teachers with various strengths of different disciplines could communicate together to make their STEM education ideas better” (teacher 1). “Some students from the disadvantaged background may find it difficult to get enough guidance, equipment, and time for STEM from their families” (teacher 4). Besides, “some schools may unwilling to give time to STEM because students are already overwhelmed with the examination pressures” (teacher 3).

4. Discussions

Our research finding suggested that the teachers’ first challenge was lack of STEM education PCK. This discovery is consistent with Geng et al.’s (2019) study in which almost half of the teacher participants thought themselves were not ready for STEM education (see also Chai et al., 2019). Providing professional development for teachers can improve their PCK, and further enhance teachers’
beliefs and attitudes, as well as easing their anxiety to adapt to the new trend (Chiu, 2017; Chiu & Churchill, 2016).

We also recommended the government to establish some public STEM centers for the teachers to build connections with other STEM educators, as teachers’ community of practice will be beneficial for them to learn from each other and design better lesson plans (Jong, 2019; Jong et al., 2008, 2010). Providing teachers with more opportunities to collaborate with others to plan their lessons (Shernoff et al., 2017) was mentioned in the previous research as a measure to improve teachers’ STEM teaching practice. The public STEM centers act as an external support platform for teachers.

The last suggestion was the public should realize the value of STEM education and embrace it in school education. Public awareness is a critical external support for teachers to implement STEM education. The STEM centers should welcome the public to experience schools’ STEM works and allow them to better understand the impact of STEM education on students and society.

5. Conclusion

Overall, the two major implementation challenges of STEM education identified in this research were on teachers’ PCK and external support. We suggested providing professional development activities focusing on PCK, building public STEM centers, and arousing public awareness to support teachers to overcome the challenges. Regarding the research limitations, for example, all the interviewees were male teachers with more than five years of teaching experience, weakening the representativeness of the research samples. It was due to the low acceptance rate of the interview invitation, at 8%. Future researchers should use a purposeful sampling method or expand the invitation pool to have more diverse representativeness.

References


Emergency remote teaching in low-resource contexts: How did teachers adapt?

Victoria ABOU-KHALIL*, Samar HELOU, Eliane KHALIFÉ, Rwitajit MAJUMDAR & Hiroaki OGATA

*Academic Center for Computing and Media Studies, Kyoto University, Japan
bGlobal Center for Medical Engineering and Informatics, Osaka University, Japan
cFaculty of Management, Saint Joseph University, Lebanon
*daboukhalil@gmail.com

Abstract: The COVID-19 pandemic forced students and teachers to engage in emergency remote learning. Remote learning is particularly challenging for students and teachers in low-resource contexts. We aim to analyze the adaptation process of teachers who engaged in remote emergency teaching in low-resource contexts. We conduct ten interviews with teachers who provided emergency online teaching in Lebanon. We show that there was a decrease in synchronous student-teacher, student-content, and student-student interactions due to the scarcity of resources. We also show how the teachers adapted their practice to cope with these challenges, and how their adaptation process increased asynchronous student-teacher and student-content interactions.

Keywords: emergency remote teaching, low-resource, COVID-19, adaptation

1. Introduction

The COVID-19 pandemic forced teachers and students into a sudden transition to emergency online learning without prior preparation or guidelines. Recent reports highlighted the challenges of this transition and exposed a significant gap in teachers’ readiness to use technology for remote teaching (Trust & Whalen, 2020). This transition has been particularly challenging and frustrating for teachers in developing countries or rural areas that have access to limited resources. In fact, teachers and students in low-resource contexts suffer from low internet connectivity, low Technological Pedagogical Content Knowledge (TPACK) (Mishra & Koehler, 2006), limited tools and financial support— which hinder their use of technologies to engage in remote education (Dalal et al., 2017; Mnyanyi & Mbwette, 2009). To deepen the understanding of the transition to emergency remote teaching in low-resource contexts, we aim to analyze the adaptation process of teachers practicing in those contexts during the COVID-19 pandemic. Our study uses Moore (1989)’s framework as an analytical framework for categorizing the effects of low resources on the adaptation process. Moore’s framework proposes three interaction categories that are essential to online learning: student–student, student–teacher, and student–content interactions.

The purpose of this study is to understand the challenges that teachers in low-resource contexts faced when transitioning to emergency remote teaching, how they adapted their practice to cope with these challenges, and how the adaptation process affected the transactional relationships among students, teachers, and content.

2. Methods

We conducted ten structured interviews with university teachers in Lebanon. The participants were purposively selected to obtain a maximum variation sample in terms of teachers’ characteristics (gender, institution, courses handled, and location). The interviews explored three questions: (1) the challenges that teachers face, (2) the strategies they employ, and (3) their general perception of emergency remote teaching. The interviews were recorded and transcribed maintaining anonymity. A thematic analysis was carried out following the guide of Braun & Clarke (Braun & Clarke, 2006). After drafting the coding schemes, we identified the themes and divided them into 1) resource-related issues 2) the related challenges that teachers face, 3) which student-teacher-content interactions are affected, 4) the strategies employed to deal with the challenges, and 5) the effect of the adaptation process on student-teacher-content interactions.
3. Results

The adaptation process of the teachers is shown in Figure 1. Two main resource-related issues were identified: low internet connectivity and lack of TPACK. These limitations imposed a decrease in three synchronous student-teacher-content interactions. To counter this decrease, the teachers adapted strategies that increased the asynchronous student-teacher and student-content interactions. However, the decrease in student-student interaction was not addressed.

![Diagram of interaction changes]

**Figure 1.** How teachers adapted to emergency remote teaching in a low-resource context

**Decrease of student-teacher synchronous interaction:** Due to bandwidth limitations, the teachers encouraged the students to disable their video when using videoconferencing platforms. This led to a decrease in audio-visual feedback from the students and consequently to a decrease in student-teacher synchronous interactions. Teachers reported having the feeling that they are talking to themselves and that they could not assess the engagement and comprehension of the students while giving their lecture.

**Increase of student-teacher asynchronous interaction:** To compensate the decrease in student-teacher synchronous interaction, the teachers increased the asynchronous student-teacher interaction. Teachers reported sharing their personal contact information with the students to provide them with extra support. This required the teachers to extend their availability and provide support during non-working hours.

**Decrease of student-content synchronous interaction:** Frequent connection failures, digital lags, delays and echoes, reduced the content delivered through the lecture as teachers tended to speak slower and often had to repeat themselves. Due to lack of interactivity during the online class, we considered that the activity of watching the lectures online fall under the student-content interaction. Moreover, due to a lack of TPACK, teachers were unable to adapt all their usual classroom content and deliver it remotely. The contents that are usually provided during the class like group case solving, or readings were removed from the synchronous classes because teachers could not supervise the work.

**Increase of student-content asynchronous interaction:** To counter the decrease in synchronous student-content interaction, teachers shared additional resources on the Learning Management System.

**Decrease of student-student synchronous interaction:** Due to the lack of TPACK, the teachers had a limited knowledge of online tools and features that allow synchronous group work between students. Consequently, student-student synchronous interaction decreased during emergency remote learning.

4. Conclusions

Our study showed that the main challenges that teachers faced when providing remote emergency teaching in low-resource contexts resulted from low internet connectivity and lack of TPACK. These challenges translated into a decrease in synchronous student-teacher, student-content, and student-student interactions. The strategies that the teachers employed to cope with these challenges increased asynchronous student-content interactions and student-teacher interaction, which is the most important interaction for students’ engagement (Martin & Bolliger, 2018). However, the student-student interaction could not be salvaged even though it can increase students’ engagement and reduce the sense...
of isolation (Abrami et al., 2011; Banna et al., 2015). Our findings align with recent work that showed that despite teachers’ unpreparedness for emergency remote teaching, they somehow managed to make online learning work (Gudmundsdottir & Hathaway, 2020). The rapid adaptation may have been catalyzed by the previous experiences of the teachers who are used to continuous crisis marked by rapid change and uncertainties (Khalifé & de Montmorillon, 2018) and their resilience and ability to harness contextual resources to effectively navigate through emerging challenges (Beltman, 2015).

To accommodate teachers’ transition to emergency online teaching in low-resource contexts, further research is needed to understand the situated needs of the teachers. These situated needs can be identified by looking at how teachers appropriated the existing technologies and used them in ways that were not intended by the original designers (Carroll, 2004; Helou et al., 2019). Moreover, past research showed that in low-resource communities, the socio-economic situation of the student majorly affects their access to ICT tools and to a positive learning environment at home (Abou-Khalil et al., 2019). To help teachers accommodate their students’ transition to emergency online learning, equity and poverty education needs to be integrated into the teachers’ professional development programs (Rowan et al., 2019). Finally, educational planning in emergencies needs to build a functional remote educational system that allows the teachers, independently of resource availability, to foster all types of transactional interaction between themselves, their students, and their content.

Acknowledgements
We would like to thank all the teachers that participated accepted to be interviewed. The research is funded by the Japan Society of Promotion of Science (JSPS) KAKENHI Grant Number JP12K34567.

References


Teacher Professional Development based on the 
DigCompEdu Framework

Mina GHOMIa* & Niels PINKWARTa
aDepartment of Computer Science, Humboldt-Universität zu Berlin, Germany
*mina.ghomi@hu-berlin.de

Abstract:
The European Framework for the Digital Competence of Educators (DigCompEdu) describes in six competence areas, which digital competences educators need specific to their profession in order to be able to integrate digital technologies in education. By referring to DigCompEdu we examine how a teacher professional development (TPD) course should be designed to promote them. Current research results on effective TPD courses, were used as a basis for the design of TPD courses to promote DigCompEdu competences. In three iterations the TPD concept was analysed, conducted and further developed. The last iteration of the execution was paused due to Covid-19. Preliminary results suggest that the TPD supports teachers in improving their digital competence. A final analysis of the data is expected in early 2021.

Keywords: Teacher Professional Development, DigCompEdu, Digital Competence

1. Introduction

Digital learning tools (e.g. tools for online collaboration, learning management systems) are not yet used by teachers in all countries (Fraillon et al., 2020). The use of information and communications technologies (ICT) by teachers in the classroom is limited in frequency and complexity (Fraillon et al., 2020). There are significant differences between countries in the expectations and requirements for teachers' qualifications for the use of ICT and for teachers' participation in professional learning activities related to ICT use (Fraillon et al., 2020). It is up to teachers not only to promote the digital competence of their students, but also to integrate digital technologies into their teaching in order to improve and innovate it. The European Framework for the Digital Competence of Educators (DigCompEdu) describes which profession-specific digital competences teachers need.

In this article we examine how a teacher professional development (TPD) course should be designed to promote the DigCompEdu competences. After a short introduction of the DigCompEdu framework, design principles for TPD courses derived from literature are presented. The resulting TPD program is currently being conducted and evaluated in Germany. First results will be presented.


DigCompEdu describes the digital competence of educators at all levels of education in six competence areas (Redecker, 2017). In this article we focus on the digital competence of secondary education teachers. The first area includes competences concerning the use of digital technologies for professional communication, collaboration and development. For example, a teacher should choose an appropriate communication tool according to the occasion. The second area describes competences for preparing lessons and deals with the selection, modification and publication of digital resources and the associated legal requirements. Hence, teachers should be familiar with copyright and data protection. The implementation and orchestration of digital technologies in teaching and supporting self-regulated learning processes as well as collaborative learning activities is addressed by the third area. The fourth area focuses not only on the collection of data through formative and summative assessments, but in particular on their critical analysis and the resulting adaptation of teaching strategies and feedback. According to the fifth competence area, teachers should be able to use digital technologies to support learner-centred pedagogical approaches. In addition to learner activation and differentiation, this also includes ensuring accessibility for all students. Pursuant to competence area six, it is also one of the
teacher's responsibilities to actively promote the digital competence of students by fostering learners’ information and media literacy and their digital communication and collaboration skills.

Each DigCompEdu competence is described in six proficiency levels (A1, A2, B1, B2, C1, C2) with a cumulative progression. A-level teachers have little or no experience in teaching with digital technologies. B-level teachers integrate different technologies in a variety of ways and contexts. C-level teachers are role models for colleagues and experiment and reflect on new innovative approaches.

Based on DigCompEdu and its progression model, a self-assessment tool (SAT) was developed and piloted with 335 participants (Ghomi & Redecker, 2019). The results suggest that the SAT is a reliable and valid instrument to measure teachers’ digital competence. However, teachers can use the tool to identify their weaknesses and determine training needs. In order to empower teachers to use digital technologies in class, TPD courses are an appropriate instrument (Lipowsky & Rzejak, 2015).

3. Effective Teacher Professional Development

Lipowsky and Rzejak (2015) and Darling-Hammond et al. (2017) summarized results of research on TPD and identified several characteristics of effective TPDs, which will be briefly outlined here. In order to develop competencies, TPD participants need time and opportunities to apply new knowledge, deal with deeper learning tasks and reflect on their experiences. It is not possible to specify a specific length of training courses, but a one-time and short workshop is not recommended for the further development of competencies. It is therefore not surprising that research on effective TPDs often examines training series that are planned over several days and a longer period of time, where input, practice and reflection phases are combined. This also allows to promote long-term cooperation between the teachers and to initiate peer learning, thus creating a professional learning community. In the time between the courses, new methods and knowledge can be practised in class and reported and reflected on in the following session. In all phases, there should always be a focus on the mostly subject-specific learning process of the students and how this can be improved. Furthermore, the results of research into the effective design of school teaching (e.g. effective classroom management, clarity and structure of instruction) also highly applicable to TPD teaching.

Based on DigCompEdu and the characteristics for effective TPDs, a TPD program consisting of four one-day events in intervals of at least 3-4 weeks was developed for secondary school teachers in Germany. In a design-based research approach, a first a theory-based TPD concept was piloted with 23 teachers in 2018 (Wang & Hannafin, 2005). By analysing the feedback surveys and observation sheets, the concept was further developed and evaluated once again with five teacher groups (N2 = 72) in 2019. The following table shows the design principles derived from it, which were now also used for the third iteration of implementation with eight teacher groups (N3 = 150). Due to school shutdowns during the Covid-19 pandemic, the study is paused for the time being. It is planned to complete all TPD courses in winter 2020 and finally to evaluate the entire data collected in 2021.

Table 1. TPD Design principles to promote DigCompEdu

<table>
<thead>
<tr>
<th>TPD Design Principles for each DigCompEdu Competence Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Professional Engagement</td>
</tr>
<tr>
<td>• Systematically use tools for communication and collaboration during and for preparation and follow-up of the TPD, e.g. discussion forum, experience blog</td>
</tr>
<tr>
<td>• Initiate exchange of experience and reflection</td>
</tr>
<tr>
<td>• Integrate online courses and materials, e.g. as a blended learning concept</td>
</tr>
<tr>
<td>2. Digital Resources</td>
</tr>
<tr>
<td>• Present a variety of tools for the creation of digital resources of different types</td>
</tr>
<tr>
<td>• Allow teachers to try out and test tools and discuss the benefits</td>
</tr>
<tr>
<td>• Show several websites for Open Educational Resources and let teachers rate them</td>
</tr>
<tr>
<td>• License materials, show how to license properly and sensitise to copyright</td>
</tr>
<tr>
<td>• Protect teachers' data and sensitise to data protection</td>
</tr>
<tr>
<td>3. Teaching and Learning</td>
</tr>
<tr>
<td>• Use methods that they will later use as teachers in class</td>
</tr>
<tr>
<td>• Be a role model in dealing with technical problems</td>
</tr>
<tr>
<td>• Let teachers develop collaborative teaching concepts</td>
</tr>
<tr>
<td>• Arrange group work online collaboratively</td>
</tr>
<tr>
<td>• Offer a choice of topics and let teachers choose which ones they want to work on</td>
</tr>
</tbody>
</table>
• Challenge teachers to implement new content and methods directly in their class

4. Assessment
• Use a variety of formative and summative assessments in appropriate situations
• Use interactive exercises when working on CPD topics with automated feedback
• Analyse the generated data and provide targeted feedback to teachers
• Discuss together the benefits and limitations of assessments
• Obtain feedback from participants using digital technologies

5. Empowering Learners
• Ensure that all they have access to materials and meet the technical requirements
• Identify the learning needs of the teachers and design the lessons so that everyone can achieve their own learning goals at their own pace and level
• Integrate hands-on activities in which teachers are expected to work creatively and actively with a subject matter

6. Facilitating Learners’ Digital Competence
Integrate projects (if possible over a longer period):
• have teachers research a topic they are not yet familiar with
• let them create digital content in compliance with copyright and privacy laws
• encourage them to communicate and collaborate online after the course

To evaluate the design principles and the effectiveness of the TPD, a pre- & post- as well as follow-up online survey was or will be sent to all participants. In addition to the IT equipment, type of use and frequency of use, the main questions asked are attitude, motivation, media-related self-efficacy, the DigCompEdu self-assessment and experience with TPDs. After each event, two teachers were interviewed and all participants were asked for feedback via an online survey on the respective event.

Related to the DigCompEdu self-assessment of the participants, first results of the conduct in 2019 suggest that more teachers reached DigCompEdu levels B1, B2 and C1 after the program. Especially in the four competence areas 2, 3, 4 and 5, the participants achieved on average more points and thus a higher level after the training. Preliminary results of the interviews and the feedback indicate that teachers at A1 level in particular need additional support, e. g. in the form of a technical briefing and step-by-step instructions as there is often a lack of everyday experiences with digital technologies.

4. Conclusion
In order to use the potential of digital technologies to improve teaching and learning, and to promote digital competence among students, teachers themselves need to develop their profession-specific digital competence as described in the European Framework DigCompEdu. As an outcome of a design-based research approach with three iterations of analysis, design, development and implementation, a proposal for a design of an effective DigCompEdu-TPD is presented in Table 1. Preliminary results suggest that the program supports teachers in improving their digital competence. A final analysis of the data and evaluation of the TPD is expected in 2021.

References
Bridging School-based Formal and Informal Learning Spaces: A Case Study of Advancing Interest-driven Education in Singapore Schools

Longkai WU*, Sujin HE
National Institute of Education, Nanyang Technological University, Singapore
*longkai.wu@nie.edu.sg

Abstract: An interest-driven learning environment can be viewed as an entry point for learning staged in a specific context, which can be formal and/or informal, with supporting and meaningful learning that engages students’ interest (Zosh, Fisher, Golinkoff & Hirsh-Pasek). This suggests the need of creating activities to develop and connect students’ understanding of their surrounding world (e.g. students’ school curriculum and social world) to what they learn in their classrooms (Samuelsson, 2005). Our study investigated how an interest-driven learning environment can serve as a platform for students to establish their interests within the formal setting of their school environment – to learn, to develop their individual agency, and to develop the knowledge and skills to accomplish their interest-driven goals.

Keywords: Interest-driven Learning, Formal and Informal Learning

1. Introduction

In Singapore, the formal classroom is recognised as the locus where learning takes place and where most activities are intentionally planned to meet curricular goals and outcomes. However, it is estimated that during their schooling years 86.7% of students’ time will be spent outside of a classroom (Gerber, Cavallo, & Edmund, 2001). Currently, little prominence has been given to the value of learning based on students’ interests, especially through an informal setting. In fact, there is often tension when informal learning sites are viewed in relation to traditional classroom spaces, as parents and even some educators question how such experiences may be beneficial to students. There are learning opportunities found in both the informal and formal environments, but there is a lack of synchronicity between the two environments. Learning (and not learning) in these two spaces are totally isolated from each other. Our aim is to make learning occur in both the formal and informal settings in an orchestrated manner, by amplifying interest through the informal approach into the formal learning space.

2. The Study

Our study investigated how an interest-driven learning environment can serve as a platform for students to establish their interests within the formal setting of their school environment – to learn, to develop their individual agency, and to develop the knowledge and skills to accomplish their interest-driven goals.

An interest-driven learning environment can be viewed as an entry point for learning staged in a specific context, which can be formal and/or informal, with supporting and meaningful learning that engages students’ interest (Zosh, Fisher, Golinkoff & Hirsh-Pasek). This suggests the need of creating activities to develop and connect students’ understanding of their surrounding world (e.g. students’ school curriculum and social world) to what they learn in their classrooms (Samuelsson, 2005). In this study, we enacted a series of Science lessons anchored in student interests in a primary school. Lessons were conducted on the topics of light and heat. We first determined the general interests that
students had, through a survey, and categorised them into four main groups (i) performing arts, (ii) arts & crafts, (iii) sports and (iv) games & leisure. The classification into four interest groups, which we later termed it, was adapted from the four distinct categories of school-based co-curricular activities (CCAs) where choices of participation in the latter is often made by students, rather than imposed on them. At the end of each of the two topics, students were given some time to work on tasks based on the four interest groups. We posit that in using student interest as examples to link to theoretical concepts, rather than just dealing with the content and outcomes of learning, could be one way to build in students’ genuine interest in learning. Therefore, with that, we can bridge student interest, garnered from their interests and incorporate that into formal classrooms learning and vice versa. This can be helpful in developing understanding of scientific concepts which are commonly seen as abstract and inaccessible to students. We collected and analysed lesson observation notes, student surveys, students’ SA1/SA2 results, and Focus Group Discussions (FGD) that were conducted with both students and teachers.

Based on our findings, it appears that by capitalising on and catering to students’ varying abilities and interests, a learning environment can pave the way for more opportunities for students to learn that is driven by their interests, agency, and relevancy. In doing so, students can develop skills and competencies that go beyond routinized cognitive tasks, such as the ability to critically seek and synthesise information, the ability to create and innovate, and the ability to self-direct one’s learning (Dede 2010).

An impactful outcome for the study was the two lesson packages that comprised the four main categories of students’ interests, and the learning frameworks for its implementation. Student interest categories guided the design and implementation of the lesson packages. The teachers, together with the researchers, were actively involved in formulating design principles to guide the implementation of the lessons. Teachers’ awareness of intentional design and facilitation in adapting to the profiles of students’ interests for positive cognitive performance to occur is deemed vital.

3. Conclusion

This study has provided an example of how interest-driven learning can be enacted in schools, with alignment to the curriculum. This is done by synchronising learning between the formal and informal settings, by amplifying learners’ interest through the informal approach within the formal learning space. Capitalising on students’ interest helped them develop skills and competencies that go beyond routinized cognitive tasks, and supported their ability to critically seek and synthesise information that often comes from outside of the formal curriculum.

References

INDEX

A
ABADIA, Rhoda 47
ABADIA, Rhodora 670
ABDELAAL, Shokry 670
ABOU-KHALIL, Victoria 686
ABURATANI, Tomoki 93
ADJEI, Seth 622
AGAPITO, Jenily L. 370
AGRAWAL, Harshit 99
AHMED, Ali 376, 664
AKÇAPINAR, Gökhan 342, 474
AKCHAPINAR, Gokhan 516
ALI, Mohommed Eunus 225
AWWAL, Nafisa 109
AYABE, Kousei 255
ABADIA, Rhoda 47

B
BADAR, Faisal 388
BAGHAEI, Nilufar 376, 436, 664
BAKER, Ryan S. 195, 622
BAKHARIA, Aneesha 215
BAKILAPADAVU, Geetha 127
BALAN NAIR, Mahmoud Mohamed Hus 516
BAUTISTA, Aaron Nol 183
BERAQUIT, Jose Isidro 370
BISWAS, Gautam 137
BOONRAT, Wisanoo 376
BOW, Catherine 546
BRETANA, Neil Arvin 47
BRINGULA, Rex 403

C
CAI, Su 360
CAI, Yuyang 552
CAI, Yuying 119
CAO, Miao 504
CASEY, John 376
CATEDRILLA, Jypzie 87, 351
<table>
<thead>
<tr>
<th>Name</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHAI, Ching Sing</td>
<td>41</td>
</tr>
<tr>
<td>CHAN, Shiau-Wei</td>
<td>171</td>
</tr>
<tr>
<td>CHANDRASEKHARAN, Sanjay</td>
<td>99</td>
</tr>
<tr>
<td>CHEANG, Gerald</td>
<td>47</td>
</tr>
<tr>
<td>CHEN, Fan</td>
<td>186</td>
</tr>
<tr>
<td>CHEN, Feixiong</td>
<td>582</td>
</tr>
<tr>
<td>CHEN, Irene Y.L.</td>
<td>267, 284, 289, 307</td>
</tr>
<tr>
<td>CHEN, Jun-Ming</td>
<td>382, 442</td>
</tr>
<tr>
<td>CHEN, Mei-Rong Alice</td>
<td>127, 235, 342, 474, 484, 680</td>
</tr>
<tr>
<td>CHEN, Wenli</td>
<td>165</td>
</tr>
<tr>
<td>CHEN, Zhi-Hong</td>
<td>448, 472</td>
</tr>
<tr>
<td>CHENG, Kun-Hung</td>
<td>348</td>
</tr>
<tr>
<td>CHENG, Lu</td>
<td>186</td>
</tr>
<tr>
<td>CHENG, Wen-Wen</td>
<td>522, 573</td>
</tr>
<tr>
<td>CHEOK, Mei Lick</td>
<td>640</td>
</tr>
<tr>
<td>CHING, Michelle Renee</td>
<td>87</td>
</tr>
<tr>
<td>CHIU, Ming Ming</td>
<td>494</td>
</tr>
<tr>
<td>CHIU, Thomas</td>
<td>41</td>
</tr>
<tr>
<td>CHIU, Thomas K.F.</td>
<td>683</td>
</tr>
<tr>
<td>CHUA, Caslon</td>
<td>225</td>
</tr>
<tr>
<td>CORRIN, Linda</td>
<td>215</td>
</tr>
<tr>
<td>CU, Jocelynn</td>
<td>183</td>
</tr>
<tr>
<td>DAI, Yun</td>
<td>41</td>
</tr>
<tr>
<td>DE LA CUESTA, Josephine</td>
<td>87, 351</td>
</tr>
<tr>
<td>DE-TORRE-BARRIO, Lucas</td>
<td>57</td>
</tr>
<tr>
<td>DEMESHKANT, Nataliia</td>
<td>591</td>
</tr>
<tr>
<td>DÍAZ-RIBAGORDA, Óscar</td>
<td>57</td>
</tr>
<tr>
<td>DIMITROVA, Vania</td>
<td>11</td>
</tr>
<tr>
<td>DONG, Yan</td>
<td>504</td>
</tr>
<tr>
<td>DUNN, Jonathan</td>
<td>1</td>
</tr>
<tr>
<td>EBARDO, Ryan</td>
<td>87, 351</td>
</tr>
<tr>
<td>EL SEBAALY, Joe</td>
<td>622</td>
</tr>
<tr>
<td>ENOMOTO, Hiroyoshi</td>
<td>75</td>
</tr>
<tr>
<td>FAN, Yang-Hsin</td>
<td>472</td>
</tr>
<tr>
<td>FARZINNEJAD, Forouzan</td>
<td>436</td>
</tr>
<tr>
<td>FLANAGAN, Brendan</td>
<td>127, 235, 261, 284, 289, 342, 474, 634, 680</td>
</tr>
</tbody>
</table>
FLOOD, Marebu 388
FORTE, Shanize 99
FUERTEZ, Enrique 466

G
GARCIA, Jaime 540
GERARDO, Jabin Raymond 183
GHOMI, Mina 689
GONZALEZ, Manuel 466
GORHAM, Tom 570
GROMIK, Nicolas 646
GUERRERO-GARCÍA-HERAS, Claudia 57
GUIJARRO-MATA-GARCÍA, María 57

H
HAKAMATA, Hiroki 561
HÄKKINEN, Päivi 109
HASNINE, Mohammad Nehal 342, 516
HAYASHI, Yugo 119, 177
HAYASHI, Yuki 63, 93
HAYASHI, Yusuke 21, 75, 189
HE, Sujin 692
HELOU, Samar 686
HERRAS, Ingrid Yvonne 370
HEW, Khe Foon 69, 612
HIRASHIMA, Tsukasa 21, 75, 189
HO, Weng Kin 171
HONG, Helen 159
HSU, Hui-Ling 448
HSU, Tien-Yu 382, 442
HU, Jon-Fan 522
HUANG, Anna Y.Q. 267, 307
HUANG, Chi-Fang 472
HUANG, Weijiao 69
HUANG, Wendy 171
HUANG, Zheng 567
HUNG, Chih-Te 558
HUTCHINS, Nicole 137
HWANG, Gwo-Jen 474, 484
<table>
<thead>
<tr>
<th>Name</th>
<th>Scroll</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>IKEDA, Mitsuru</td>
<td></td>
<td>93</td>
</tr>
<tr>
<td>ITOH, Yukihiro</td>
<td></td>
<td>205, 255, 561, 658</td>
</tr>
<tr>
<td>JAYARATNA, Sameera</td>
<td></td>
<td>225</td>
</tr>
<tr>
<td>JAYAWICKRAMA, Thilini</td>
<td></td>
<td>670</td>
</tr>
<tr>
<td>JIAO, Xinyue</td>
<td></td>
<td>360</td>
</tr>
<tr>
<td>JONG, Bin-Shyan</td>
<td></td>
<td>267</td>
</tr>
<tr>
<td>JONG, Morris S. Y.</td>
<td></td>
<td>683</td>
</tr>
<tr>
<td>JONG, Morris Siu-Yung</td>
<td></td>
<td>41</td>
</tr>
<tr>
<td>KANG, Ching-Hsiang</td>
<td></td>
<td>267</td>
</tr>
<tr>
<td>KANNAN, Vijayanandhini</td>
<td></td>
<td>602</td>
</tr>
<tr>
<td>KAPP, Tina</td>
<td></td>
<td>622</td>
</tr>
<tr>
<td>KARNAM, DurgaPrasad</td>
<td></td>
<td>99</td>
</tr>
<tr>
<td>KASHIHARA, Akihiro</td>
<td></td>
<td>312</td>
</tr>
<tr>
<td>KASIRI, Mohsen</td>
<td></td>
<td>436</td>
</tr>
<tr>
<td>KAUSHAL, Kritika</td>
<td></td>
<td>47</td>
</tr>
<tr>
<td>KHALIFÉ, Eliane</td>
<td></td>
<td>686</td>
</tr>
<tr>
<td>KHAN, Khalid</td>
<td></td>
<td>646</td>
</tr>
<tr>
<td>KHATRI, Shreya</td>
<td></td>
<td>47</td>
</tr>
<tr>
<td>KHEZRIAN, Hadi</td>
<td></td>
<td>436</td>
</tr>
<tr>
<td>KIM, Mi Song</td>
<td></td>
<td>171</td>
</tr>
<tr>
<td>KIRI, Hinal</td>
<td></td>
<td>99</td>
</tr>
<tr>
<td>KOGURE, Satoru</td>
<td></td>
<td>205, 255, 561, 658</td>
</tr>
<tr>
<td>KOH, Elizabeth</td>
<td></td>
<td>159</td>
</tr>
<tr>
<td>KOJIRI, Tomoko</td>
<td></td>
<td>63</td>
</tr>
<tr>
<td>KONDO, Makoto</td>
<td></td>
<td>561</td>
</tr>
<tr>
<td>KONDO, Taisho</td>
<td></td>
<td>394</td>
</tr>
<tr>
<td>KONDO, Taisyo</td>
<td></td>
<td>272</td>
</tr>
<tr>
<td>KONG, Siu-Cheung</td>
<td></td>
<td>652</td>
</tr>
<tr>
<td>KONISHI, Tatsuhiro</td>
<td></td>
<td>205, 255, 295, 561, 658</td>
</tr>
<tr>
<td>KUO, Che-Yu</td>
<td></td>
<td>307</td>
</tr>
<tr>
<td>KUROMIYA, Hiroyuki</td>
<td></td>
<td>272</td>
</tr>
<tr>
<td>KUTAY, Cat</td>
<td></td>
<td>540</td>
</tr>
<tr>
<td>LAI, Ming</td>
<td></td>
<td>652</td>
</tr>
<tr>
<td>LALLAVE, Harvey</td>
<td></td>
<td>183</td>
</tr>
<tr>
<td>Name</td>
<td>Page(s)</td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>LAM, Sze Ming</td>
<td>494</td>
<td></td>
</tr>
<tr>
<td>LAPINID, Minie Rose</td>
<td>183</td>
<td></td>
</tr>
<tr>
<td>LATORRE, Patrick Luigi</td>
<td>183</td>
<td></td>
</tr>
<tr>
<td>LEÑO, Cecilia</td>
<td>87</td>
<td></td>
</tr>
<tr>
<td>LECAILLIEZ, Louis</td>
<td>261</td>
<td></td>
</tr>
<tr>
<td>LEE, Chien-Sing</td>
<td>147</td>
<td></td>
</tr>
<tr>
<td>LEE, Lung-Hao</td>
<td>558</td>
<td></td>
</tr>
<tr>
<td>LEI, Chunlin</td>
<td>552</td>
<td></td>
</tr>
<tr>
<td>LI, Baoping</td>
<td>318</td>
<td></td>
</tr>
<tr>
<td>LI, Huiyong</td>
<td>330, 394</td>
<td></td>
</tr>
<tr>
<td>LI, Luyue</td>
<td>153</td>
<td></td>
</tr>
<tr>
<td>LIANG, Hsin-Yi</td>
<td>382, 442</td>
<td></td>
</tr>
<tr>
<td>LIMJAP, Auxencia</td>
<td>183</td>
<td></td>
</tr>
<tr>
<td>LIMPIN, Laiza</td>
<td>87</td>
<td></td>
</tr>
<tr>
<td>LIN, Pei-Yi</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>LIN, Po-Chen</td>
<td>558</td>
<td></td>
</tr>
<tr>
<td>LIN, Yi-Hsuan</td>
<td>474</td>
<td></td>
</tr>
<tr>
<td>LIU, Bing</td>
<td>646</td>
<td></td>
</tr>
<tr>
<td>LIU, Hanyong</td>
<td>528</td>
<td></td>
</tr>
<tr>
<td>LIU, Si Hui</td>
<td>494</td>
<td></td>
</tr>
<tr>
<td>LIU, Yang</td>
<td>318, 564</td>
<td></td>
</tr>
<tr>
<td>LIU, Zhiyang</td>
<td>564</td>
<td></td>
</tr>
<tr>
<td>LIU, Zifeng</td>
<td>360</td>
<td></td>
</tr>
<tr>
<td>LOOI, Chee-Kit</td>
<td>171</td>
<td></td>
</tr>
<tr>
<td>LU, Owen H.T.</td>
<td>267, 307</td>
<td></td>
</tr>
<tr>
<td>LU, Yutong</td>
<td>454</td>
<td></td>
</tr>
<tr>
<td>LU, Zhihong</td>
<td>528</td>
<td></td>
</tr>
<tr>
<td>LUAN, Lin</td>
<td>504</td>
<td></td>
</tr>
<tr>
<td>LUNDQVIST, Karsten</td>
<td>664</td>
<td></td>
</tr>
</tbody>
</table>

**M**

<table>
<thead>
<tr>
<th>Name</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA, Ning</td>
<td>153, 186</td>
</tr>
<tr>
<td>MAHMUD, Rosaini</td>
<td>640</td>
</tr>
<tr>
<td>MAJUMDAR, Rwitajit</td>
<td>127, 235, 272, 330, 342, 394, 474, 484, 602, 680, 686</td>
</tr>
<tr>
<td>MAJUMDER, Reek</td>
<td>127</td>
</tr>
<tr>
<td>MAKHLOUF, Jihed</td>
<td>336</td>
</tr>
<tr>
<td>MANAHAN, Dominique Marie Antoinette</td>
<td>370</td>
</tr>
<tr>
<td>MANERO, Borja</td>
<td>57, 466</td>
</tr>
<tr>
<td>MARTINS, Rafael M.</td>
<td>278</td>
</tr>
<tr>
<td>MASIELLO, Italo</td>
<td>278</td>
</tr>
<tr>
<td>MASON, Jon</td>
<td>388, 646</td>
</tr>
<tr>
<td>MCLAREN, Bruce M.</td>
<td>195</td>
</tr>
<tr>
<td>Name</td>
<td>Page Numbers</td>
</tr>
<tr>
<td>---------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>MD. KHAMBARI, Mas Nida</td>
<td>510</td>
</tr>
<tr>
<td>MILRAD, Marcelo</td>
<td>278</td>
</tr>
<tr>
<td>MINE, Tsunernori</td>
<td>336</td>
</tr>
<tr>
<td>MINEMATSU, Tsubasa</td>
<td>301</td>
</tr>
<tr>
<td>MISHRA, Shitanshu</td>
<td>137</td>
</tr>
<tr>
<td>MITROVIC, Antonija</td>
<td>1, 416</td>
</tr>
<tr>
<td>MITSUHARA, Hiroyuki</td>
<td>400</td>
</tr>
<tr>
<td>MOHAMED RAZALI, Abu Bakar</td>
<td>510</td>
</tr>
<tr>
<td>MOHAMMADHASSAN, Negar</td>
<td>1</td>
</tr>
<tr>
<td>MOHAMMED, Abrar</td>
<td>11</td>
</tr>
<tr>
<td>MOHD AYUB, Ahmad Fauzi</td>
<td>640</td>
</tr>
<tr>
<td>MOHSENI, Zeynab</td>
<td>278</td>
</tr>
<tr>
<td>MONTES, Jagoba</td>
<td>466</td>
</tr>
<tr>
<td>MOORE, Steven</td>
<td>31</td>
</tr>
<tr>
<td>MORA, Kevin Arnel C.</td>
<td>370</td>
</tr>
<tr>
<td>MORENO, Ma. Monica L.</td>
<td>370</td>
</tr>
<tr>
<td>MOURI, Kousuke</td>
<td>516, 534</td>
</tr>
<tr>
<td>MURAMATSU, Masaya</td>
<td>205</td>
</tr>
<tr>
<td>MURATA, Ryusuke</td>
<td>301</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>NADAF, Ali</td>
<td>622</td>
</tr>
<tr>
<td>NAIR, Madhavan BALAN</td>
<td>147</td>
</tr>
<tr>
<td>NAIR, Sidarth</td>
<td>47</td>
</tr>
<tr>
<td>NAKANISHI, Taro</td>
<td>272</td>
</tr>
<tr>
<td>NAVARRO, Javier</td>
<td>466</td>
</tr>
<tr>
<td>NESHATIAN, Kourosh</td>
<td>1</td>
</tr>
<tr>
<td>NG, Betsy</td>
<td>159</td>
</tr>
<tr>
<td>NGUYEN, Huy</td>
<td>31</td>
</tr>
<tr>
<td>NIU, Jiayu</td>
<td>186</td>
</tr>
<tr>
<td>NOGUCHI, Yasuhiro</td>
<td>205, 255, 561, 658</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>OBEODOZA, Jarl Brent</td>
<td>426</td>
</tr>
<tr>
<td>OCHEJA, Patrick</td>
<td>235, 261</td>
</tr>
<tr>
<td>OGATA, Hiroaki</td>
<td>127, 235, 261, 272, 284, 289, 330, 342, 394, 474, 484, 534, 570, 602, 634, 680, 686</td>
</tr>
<tr>
<td>ONG, Ethel</td>
<td>183</td>
</tr>
<tr>
<td>OOI, James Ean-Huat</td>
<td>147</td>
</tr>
<tr>
<td>OTIENO, Suzanne</td>
<td>109</td>
</tr>
<tr>
<td>OYELERE, Solomon Sunday</td>
<td>261</td>
</tr>
</tbody>
</table>
P
PAN, Qiushi 90
PANDEY, Aashi 47
PANKIEWICZ, Maciej 324
PEETERS, Ward 245
PI, Zhongling 165
PINANDITO, Aryo 189
PINKWART, Niels 689
POTYRAŁA, Katarzyna 591
PÖYSÄ-TARHONEN, Johanna 109
PUSEY, Megan 460

Q
QIN, Congyao 69

R
RAFFE, William 540
RAPPA, Natasha 460
RAWAT, Aastha 47
ROBATI, Mehdi 47
RODRIGO, Ma. Mercedes T. 370, 403

S
SAKATA, Keisuke 658
SAQR, Mohammed 245
SATO, Shusuke 406
SCRUGGS, Richard 195
SELLIS, Timos 225
SEOW, Peter 171
SEOW, Tricia 159
SETA, Kazuhisa 63, 93
SEYA, Ryotaro 312
SHARIFZADEH, Hamid 376
SHIMADA, Atsushi 301
SHIMOJO, Shigen 119, 177
SHISHIBORI, Masami 400
SISON, Raymund 426
SIU, Cora Ka-Yuk 652
SNYDER, Caitlin 137
SONG, Yanjie 534
SOTARDI, Valerie 416
<table>
<thead>
<tr>
<th>Name</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAMPER, John</td>
<td>31</td>
</tr>
<tr>
<td>SUMI, Kaoru</td>
<td>406</td>
</tr>
<tr>
<td>SUN, Daner</td>
<td>96</td>
</tr>
<tr>
<td>SURESH, Dhivya</td>
<td>159</td>
</tr>
<tr>
<td>SUZUKI, Kanako</td>
<td>295</td>
</tr>
<tr>
<td>SZAPIRO, Deborah</td>
<td>540</td>
</tr>
</tbody>
</table>

**T**

<table>
<thead>
<tr>
<th>Name</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAHIR, Faiza</td>
<td>416</td>
</tr>
<tr>
<td>TAKADA, Hideyuki</td>
<td>192</td>
</tr>
<tr>
<td>TAKEHARA, Daichi</td>
<td>81</td>
</tr>
<tr>
<td>TAKII, Kensuke</td>
<td>235, 272</td>
</tr>
<tr>
<td>TAN, Jesmine S. H.</td>
<td>165</td>
</tr>
<tr>
<td>TEE, Yi-Huan</td>
<td>159</td>
</tr>
<tr>
<td>TEZUKA, Taro</td>
<td>90</td>
</tr>
<tr>
<td>TOMCZYK, Łukasz</td>
<td>591, 628</td>
</tr>
<tr>
<td>TORRES, Johanna Marion R.</td>
<td>370</td>
</tr>
<tr>
<td>TOYOKAWA, Yuko</td>
<td>484</td>
</tr>
<tr>
<td>TRAPERO, Hazel</td>
<td>87</td>
</tr>
<tr>
<td>TSENG, Yuen-Hsien</td>
<td>558</td>
</tr>
</tbody>
</table>

**U**

<table>
<thead>
<tr>
<th>Name</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>UEDA, Hiroshi</td>
<td>516</td>
</tr>
<tr>
<td>UENO, Noboru</td>
<td>400</td>
</tr>
</tbody>
</table>

**V**

<table>
<thead>
<tr>
<th>Name</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAIDYA, Vimita</td>
<td>376</td>
</tr>
<tr>
<td>VEGAN, Nandakumara</td>
<td>147</td>
</tr>
<tr>
<td>VIBERG, Olga</td>
<td>245</td>
</tr>
</tbody>
</table>

**W**

<table>
<thead>
<tr>
<th>Name</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>WADA, Takuya</td>
<td>63</td>
</tr>
<tr>
<td>WANG, Dan</td>
<td>510</td>
</tr>
<tr>
<td>WANG, Huilin</td>
<td>528</td>
</tr>
<tr>
<td>WANG, Yuh-Shyang</td>
<td>558</td>
</tr>
<tr>
<td>WANG, Zhan</td>
<td>494</td>
</tr>
<tr>
<td>WARRIEM, Jayakrishnan</td>
<td>602</td>
</tr>
<tr>
<td>WATTERSON, Craig</td>
<td>664</td>
</tr>
<tr>
<td>WENG, Xiaojing</td>
<td>683</td>
</tr>
<tr>
<td>WIBOWO, Santoso</td>
<td>351</td>
</tr>
<tr>
<td>WONG, Kok Wai</td>
<td>460</td>
</tr>
</tbody>
</table>
WONG, Su Luan 510, 640
WU, Di 582
WU, Lan 318
WU, Longkai 171, 692

X
XU, Chang 454
XU, Xueying 186

Y
YAMAMOTO, Raiya 255
YAMAMOTO, Sho 75
YAMASHITA, Koichi 205, 255, 658
YAMMAHI, Aishah AL 622
YANG, Albert 284, 289
YANG, Bo 318
YANG, Lan 96
YANG, Stephen J.H. 267, 307
YANG, Yin 534
YANG, Yuan Yuan 394
YANG, Yuanyuan 330
YE, Fan 153
YETT, Bernard 137
YIN, Chengjiu 192
YOSHITAKE, Daichi 634
YU, Fu- Yun 522, 573
YU, Guangming 564
YU, Liqin 582
YU, Shengquan 318
YU, Xiaohan 454, 564

Z
ZHANG, Lei 153
ZHANG, Limei 567
ZHANG, Ningyu 137
ZHANG, Wanyu 153
ZHANG, Xian 454
ZHENG, Chunping 454, 528, 564
ZHOU, Juan 192
ZHU, Kaige 582
ZHU, Sha 582