Educational Technology & Society
An International Journal

Aims and Scope

Educational Technology & Society is a quarterly journal published in January, April, July and October. Educational Technology & Society seeks academic articles on the issues affecting the developers of educational systems and educators who implement and manage such systems. The articles should discuss the perspectives of both communities and their relation to each other:

- Educators aim to use technology to enhance individual learning as well as to achieve widespread education and expect the technology to blend with their individual approach to instruction. However, most educators are not fully aware of the benefits that may be obtained by proactively harnessing the available technologies and how they might be able to influence further developments through systematic feedback and suggestions.

- Educational system developers and artificial intelligence (AI) researchers are sometimes unaware of the needs and requirements of typical teachers, with a possible exception of those in the computer science domain. In transferring the notion of a “user” from the human-computer interaction studies and assigning it to the “student,” the educator’s role as the “implementer/manager/user” of the technology has been forgotten.

The aim of the journal is to help them better understand each other’s role in the overall process of education and how they may support each other. The articles should be original, unpublished, and not in consideration for publication elsewhere at the time of submission to Educational Technology & Society. The articles should be submitted electronically.

The scope of the journal is broad. Following list of topics is considered to be within the scope of the journal:


Founding Editor

Kinsbuk, University of North Texas, USA.

Editors

Nian-Shing Chen, National Sun Yat-sen University, Taiwan; Demetrios G Sampson, Curtin University, Australia.

Advisory Board

Ignacio Aedo, Universidad Carlos III de Madrid, Spain; Mohamed Ally, Athabasca University, Canada; Luis Anido-Rifon, University of Vigo, Spain; Gautam Biswas, Vanderbilt University, USA; Rosa Maria Bottino, Consiglio Nazionale delle Ricerche, Italy; Mark Bullen, University of British Columbia, Canada; Tak-Wai Chan, National Central University, Taiwan; Kuo-En Chang, National Taiwan Normal University, Taiwan; Ni Chang, Indiana University South Bend, USA; Yam San Chee, Nanyang Technological University, Singapore; Sherry Chen, Brunel University, UK; Bridget Cooper, University of Sunderland, UK; Darina Dicheva, Winston-Salem State University, USA; Jon Dron, Athabasca University, Canada; Michael Eisenberg, University of Colorado, Boulder, USA; Robert Farrell, IBM Research, USA; Brian Garner, Deakin University, Australia; Tiong Goh, Victoria University of Wellington, New Zealand; Mark D. Gross, Carnegie Mellon University, USA; Roger Hartley, Leeds University, UK; J R Isaac, National Institute of Information Technology, India; Mohamed Jenni, University of Tunis, Tunisia; Mike Joy, University of Warwick, United Kingdom; Athanasis Karoulis, Hellenic Open University, Greece; Paul Kirschner, Open University of the Netherlands, The Netherlands; William Klemm, Texas A&M University, USA; Rob Koper, Open University of the Netherlands, The Netherlands; Jimmy Ho Man Lee, The Chinese University of Hong Kong, Hong Kong; Rudy Leoluche, Universite Laval, Canada; Tzu-Chien Liu, National Central University, Taiwan; Rory McGeer, Athabasca University, Canada; David Merrill, Brigham Young University - Hawaii, USA; Marcelo Milrad, Vaxjo University, Sweden; Riihicho Mizoguchi, Osaka University, Japan; Permanand Mohan, The University of the West Indies, Trinidad and Tobago; Kiyoshi Nakabayashi, National Institute of Multimedia Education, Japan; Hiroaki Ogata, Tokushima University, Japan; Toshio Okamoto, The University of Electro-Communications, Japan; Jose A. Pino, University of Chile, Chile; Thomas C. Reeves, The University of Georgia, USA; Norbert M. Seel, Albert-Ludwigs-University of Freiburg, Germany; Timothy K. Shah, Tamkang University, Taiwan; Yoshiaki Shindo, Nippon Institute of Technology, Japan; Kevin Singley, IBM Research, USA; J Michael Spector, Florida State University, USA; Slavi Stoyanov, Open University, The Netherlands; Timothy Tso, Nanyang Technological University, Singapore; Chin-Chung Tsai, National Taiwan Normal University, Taiwan; Jie Chi Yang, National Central University, Taiwan; Stephen J. H. Yang, National Central University, Taiwan; Yu-Mei Wang, University of Alabama at Birmingham, USA; Ashok Patel, CAL Research & Software Engineering Centre, UK; Reinhard Oppermann, Fraunhofer Instituts Angewandte Informationstechnik, Germany; Vladimir A Fonichkov, K. E. Tsitskovsky Russian State Tech Univ, Russia; Olga S Fominova, Studio “Culture, Ecology, and Foreign Languages,” Russia; Piet Kommers, University of Twente, The Netherlands; Chul-Hwan Lee, Inchon National University of Education, Korea; Brent Muirhead, University of Phoenix Online, USA; Erikki Sutinen, University of Joensuu, Finland; Vladimir Uskov, Bradley University, USA.

Editorial Assistant

Sie Wai (Sylvia) Chew, National Sun Yat-sen University, Taiwan; Stelios Sergis, University of Piraeus & Centre for Research and Technology Hellas, Greece.

Technical Manager

Stelios Sergis, University of Piraeus & Centre for Research and Technology Hellas, Greece

Executive Peer-Reviewers

see http://www.j-ets.net

Publisher

National Taiwan Normal University, Taiwan

Editorial Office

c/o Dr Nian-Shing Chen, Taiwan E-Learning and Digital Content Association, National Taiwan Normal University, 129, Sec. 1, Heping E. Rd., Education Building, Rm. 611, Taipei 10664, Taiwan.
Supporting Organizations
Center for Cognition and Gesture-based Computing, National Sun Yat-sen University, Taiwan
Taiwan E-Learning and Digital Content Association (TELDCA), Taiwan
Information Technologies Institute, Centre for Research & Technology Hellas (C.R.T.H), Greece
University of Piraeus, Greece

Subscription Prices and Ordering Information
For subscription information, please contact the editors at ets.editors@gmail.com

Advertisements
Educational Technology & Society accepts advertisement of products and services of direct interest and usefulness to the readers of the journal, those involved in education and educational technology. Contact the editors at ets.editors@gmail.com

Abstracting and Indexing

Guidelines for authors
Submissions are invited in the following categories:
- Peer reviewed publications: Full length articles (4000 - 7000 words)
- Book reviews
- Software reviews
- Website reviews

All peer review publications will be refereed in double-blind review process by at least two international reviewers with expertise in the relevant subject area. Book, Software and Website Reviews will not be reviewed, but the editors reserve the right to refuse or edit review.

For detailed information on how to format your submissions, please see: https://www.etsy.net/ETS/guide.html

Submission procedure
Authors, submitting articles for a particular special issue, should send their submissions directly to the appropriate Guest Editor. Guest Editors will advise the authors regarding submission procedure for the final version.

All submissions should be in electronic form. The editors will acknowledge the receipt of submission as soon as possible.

The preferred formats for submission are Word document and RTF, but editors will try their best for other formats too. For figures, GIF and JPEG (JPG) are the preferred formats. Authors must supply separate figures in one of these formats besides embedding in text.

Please provide following details with each submission: Author(s) full name(s) including title(s), Name of corresponding author, Job title(s), Organisation(s), Full contact details of ALL authors including email address, postal address, telephone and fax numbers.

In case of difficulties, please contact ets.editors@gmail.com (Subject: Submission for Educational Technology & Society journal).
Journal of Educational Technology & Society
Volume 21 Number 2 2018

Table of contents

Full Length Articles

Multilingual Videos for MOOCs and OER
Juan Daniel Valor Miró, Pau Baquero-Arnal, Jorge Civera, Carlos Turró and Alfons Juan
1–12

A Structural Model for Students’ Adoption of Learning Management Systems: An Empirical Investigation in the Higher Education Context
Duygu Fındık-Coşkunçay, Nurcan Alkış and Sevgi Özkan-Özdemir
13–27

Instructional Suggestions Supporting Science Learning in Digital Environments Based on a Review of Eye Tracking Studies
Fang-Ying Yang, Meng-Jung Tsai, Guo-Li Chiou, Silvia Wen-Yu Lee, Cheng-Chieh Chang and Li-Ling Chen
28–45

High School Students’ Views on the PBL Activities Supported via Flipped Classroom and LEGO Practices
Baris Cakurbasi and Mubin Kiyici
46–61

Evolving Learning Paradigms: Re-Setting Baselines and Collection Methods of Information and Communication Technology in Education Statistics
David Gibson, Tania Broadley, Jill Downie and Peter Wallet
62–73

Are Games Effective Learning Tools? A Review of Educational Games
Sara de Freitas
74–84

Teaching Classical Mechanics Concepts using Visuo-haptic Simulators
Luis Neri, Julieta Noguez, Victor Robledo-Rella, David Escobar-Castillejos and Andres Gonzalez-Nucamendi
85–97

The Effects of Representation Tool (Visible-Annotation) Types to Support Knowledge Building in Computer-Supported Collaborative Learning
Yoonhee Shin, Dongsik Kim and Jaewon Jung
98–110

Modelling and Simulating Electronics Knowledge: Conceptual Understanding and Learning through Active Agency
Adrian Twissell
111–123

Representations of Animal Companions on Student Learning Perception: Static, Animated and Tangible
Zhi-Hong Chen and Sheng-Chun Wang
124–133

Guest Editorial

Digital Citizenship: Trends and Research Issues of Learning Analytics and Educational Big Data
Gwo-Jen Hwang, Daniel Spikol and Kam-Cheong Li
134–136

Special Issue Articles

Learning Behavior Analysis of a Ubiquitous Situated Reflective Learning System with Application to Life Science and Technology Teaching
Wu-Yuin Hwang, Hong-Ren Chen, Nian-Shing Chen, Li-Kai Lin and Jin-Wen Chen
137–149

Learning Analytics for Supporting Seamless Language Learning using E-book with Ubiquitous Learning System
Kousuke Mouri, Noriko Uosaki, Hiroaki Ogata
150–163

How Competition in a Game-based Science Learning Environment Influences Students’ Learning Achievement, Flow Experience, and Learning Behavioral Patterns
Ching-Huei Chen, Jun-Han Liu and Wen-Chuan Shou
164–176

A Votable Concept Mapping Approach to Promoting Students’ Attentional Behavior: An Analysis of Sequential Behavioral Patterns and Brainwave Data
Jerry Chih-Yuan Sun, Gwo-Jen Hwang, Yu-Yun Lin, Shih-Jou Yu, Liu-Cheng Pan and Ariel Yu-Zhen Chen
177–191

Using a Learner-Topic Model for Mining Learner Interests in Open Learning Environments
Pengfei Wu, Shengquan Yu and Dan Wang
192–204
<table>
<thead>
<tr>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Learning Analytics Approach to Investigating Factors Affecting EFL Students’ Oral Performance in a Flipped Classroom</td>
<td>205–219</td>
</tr>
<tr>
<td>Chi-Jen Lin and Gwo-Jen Hwang</td>
<td></td>
</tr>
<tr>
<td>Applying Learning Analytics for the Early Prediction of Students’ Academic Performance in Blended Learning</td>
<td>220–232</td>
</tr>
<tr>
<td>Owen H.T. Lu, Anna Y.Q. Huang, Jeff C. H. Huang, Albert J. Q. Lin, Hiroaki Ogata and Stephen J. H. Yang</td>
<td></td>
</tr>
<tr>
<td>Personalized Word-Learning based on Technique Feature Analysis and Learning Analytics</td>
<td>233–244</td>
</tr>
<tr>
<td>Di Zou and Haoran Xie</td>
<td></td>
</tr>
<tr>
<td>What Learning Analytics Tells Us: Group Behavior Analysis and Individual Learning Diagnosis based on Long-Term and Large-Scale Data</td>
<td>245–258</td>
</tr>
<tr>
<td>Jia-Hua Zhang, Ye-Xing Zhang, Qin Zou and Sen Huang</td>
<td></td>
</tr>
<tr>
<td>A Comparison between Two Automatic Assessment Approaches for Programming: An Empirical Study on MOOCs</td>
<td>259–272</td>
</tr>
<tr>
<td>Anis Bey, Patrick Jermann and Pierre Dillenbourg</td>
<td></td>
</tr>
<tr>
<td>Learning Analytics at Low Cost: At-risk Student Prediction with Clicker Data and Systematic Proactive Interventions</td>
<td>273–290</td>
</tr>
<tr>
<td>Samuel P. M. Choi, S.S. Lam, Kam Cheong Li and Billy T.M. Wong</td>
<td></td>
</tr>
<tr>
<td>A Peer Coaching-based Professional Development Approach to Improving the Learning Participation and Learning Design Skills of In-Service Teachers</td>
<td>291–304</td>
</tr>
<tr>
<td>Ning Ma, Shuang Xin and Jia-Yuan Du</td>
<td></td>
</tr>
<tr>
<td>Peer Assessment of Webpage Design: Behavioral Sequential Analysis Based on Eye Tracking Evidence</td>
<td>305–321</td>
</tr>
<tr>
<td>Ting-Chia Hsu, Shao-Chen Chang and Nan-Cen, Liu</td>
<td></td>
</tr>
<tr>
<td>Effects of a Progressive Prompting-based Educational Game on Second Graders’ Mathematics Learning Performance and Behavioral Patterns</td>
<td>322–334</td>
</tr>
<tr>
<td>Kai-Hsiang Yang, Hui-Chun Chu and Li-Yu Chiang</td>
<td></td>
</tr>
</tbody>
</table>
Multilingual Videos for MOOCs and OER

Juan Daniel Valor Miró1, Pau Baquero-Arnal1, Jorge Civera1*, Carlos Turró2 and Alfonso Juan1
1MLLP, DSIC, Universitat Politècnica de València, València, Spain // 2Media Services Universitat Politècnica de València, València, Spain // jvalor@dsic.upv.es // jcivera@dsic.upv.es // ajuan@dsic.upv.es // pabaar@inf.upv.es // turro@cc.upv.es

*Corresponding author

(Submitted September 30, 2016; Revised November 23, 2016; Accepted January 1, 2017)

ABSTRACT

Massive Open Online Courses (MOOCs) and Open Educational Resources (OER) are rapidly growing, but are not usually offered in multiple languages due to the lack of cost-effective solutions to translate the different objects comprising them and particularly videos. However, current state-of-the-art automatic speech recognition (ASR) and machine translation (MT) techniques have reached a level of maturity which opens the possibility of producing multilingual video subtitles of publishable quality at low cost. This work summarizes authors’ experience in exploring this possibility in two real-life case studies: a MOOC platform and a large video lecture repository. Apart from describing the systems, tools and integration components employed for such purpose, a comprehensive evaluation of the results achieved is provided in terms of quality and efficiency. More precisely, it is shown that draft multilingual subtitles produced by domain-adapted ASR/MT systems reach a level of accuracy that make them worth post-editing, instead of generating them ex novo, saving approximately 25%–75% of the time. Finally, the results reported on user multilingual data consumption reflect that multilingual subtitles have had a very positive impact in our case studies boosting student enrolment, in the case of the MOOC platform, by 70% relative.

Keywords

Video lecture repositories, MOOCs, Speech recognition, Machine translation, Multilingual

Introduction

Massive Open Online Courses (MOOCs) are rapidly growing since 2011, with more than 35 million students and 4000 courses offered at the beginning of 2016, roughly doubling the figures of the previous year (Shad 2015). Although US-based providers like edX and Coursera are now targeting international students, most courses are only delivered in English (76%), Spanish (8%), French (5%) or Chinese (3%) (see class-central.com/languages). For MOOCs to reach a worldwide audience, they should be provided in multilingual form. And this also holds true for Open Educational Resources (OER) in general. Although MOOCs and OER comprise objects of different kinds, in this work we focus our attention on producing multilingual video lectures; that is, on adding subtitles in their source (spoken) language(s) and then translate them into different target languages. Apart from its application to MOOCs and OER, multilinguality is of great interest in all contexts where educational videos are used. This includes online education in general (Kay, 2012), flipped teaching (Bishop & Verleger, 2013), and in-class recording services (Ketterl et al., 2010).

A direct approach to obtain source video subtitles is to generate automatic transcriptions by using Automatic Speech Recognition (ASR) technology. Indeed, the application of ASR technology to lecture recordings is by no means new. A detailed account of significant efforts in this domain up to 2010 can be found in (de-Pablos et al., 2011). More recent research efforts on ASR applied to educational videos can be found in the European projects transLectures (Transcription and Translation of Video Lectures) and EMMA (European Multiple MOOC Aggregator) (see platform.europeanmoocs.eu). Broadly speaking, from the results of these efforts we may conclude that ASR technology has reached a level of maturity that allows us to generate low-cost, automatic source subtitles of (nearly) publishable quality in most cases. It is worth noting, however, that such quality is only achievable by developing state-of-the-art ASR systems adapted to the particular task (media repository) at hand. In comparison with mainstream providers (e.g., YouTube), adapted systems achieve relative accuracy improvements of about 40%. In any case, even if automatic source subtitles are of moderate quality, they are often very useful for different purposes such as improving accessibility for hearing-impaired and foreign students (de-Pablos et al., 2011; Ranchal et al., 2013), video-clip search based on keywords (Repp et al., 2008) and discovery of content-related videos in a repository (Glass et al., 2007).

Analogously to the case of source subtitles, a direct approach to obtain target video subtitles is to generate automatic translations by using Machine Translation (MT) technology. This approach has been also explored with good results in transLectures and EMMA, and more recently in TraMOOC (Kordoni et al., 2016). A clear
conclusion from these results is that the translation quality of adapted MT systems is often accurate enough for post-editing; that is, it is often the case that the automatic translation is not far from the correct translation, and thus it is more time-efficient to review it than to produce the entire translation manually. In addition to this, as in ASR, system adaptation has been shown to be a key factor in maximizing output quality: in comparison with mainstream providers (e.g., Google Translate), adapted MT systems increase translation quality by about 20% relative. MT is normally applied to clean, post-edited automatic transcriptions and, as indicated above, automatic translations are also post-edited to end up with target subtitles of publishable quality. Regarding this, it is worth noting that many approaches have been considered to increase user productivity when reviewing subtitles, but post-editing is still the most popular (Plitt & Masselot, 2010; Specia, 2011; O’Brien & Simard, 2014; Valor-Miró et al., 2015).

The above discussion does not mean that the task of producing multilingual videos for MOOCs and OER simply comes down to developing advanced ASR/MT systems for lecture recordings. Obviously, it requires expertise, resources and tools from ASR/MT, but also additional components and experience for their proper integration into real-life educational environments. In this respect, this article summarizes a large part of the experience gained on this task by the Universitat Politècnica de València (UPV)’s Machine Learning and Language Processing (MLLP) group during transLectures and EMMA. Our main goal is to provide a comprehensive evaluation of the results achieved in a real-life MOOC platform and a large video lecture repository. However, resources and tools are described in broad terms since our focus here is not on ASR/MT technical details. On the contrary, here we report detailed results in terms of quality and efficiency, as well as on the impact multilingual videos have had in our real-life case studies.

The article is organized as follows. After a review of our case studies, the systems, tools and integration components required for multilingual video production are summarized. Then, detailed results on transcription and translation quality are provided, also including comparative results with mainstream providers. These results are followed by a thorough evaluation of transcription (translation) reviewing time for each language (language pair) considered separately, and also across all languages considered. Next, the impact these systems, tools and integration components have had in the case studies. Finally, the main conclusions drawn are summarized.

Case studies

This section introduces two case studies in which multilingual video subtitles are delivered: a MOOC platform and a large video lecture repository.

The EMMA platform

The European project EMMA (February 2014 – July 2016) involved 12 partners delivering more than 30 multilingual MOOCs on diverse topics.

![Figure 1. Screenshot of a trilingual MOOC](image-url)
Multilingualism is a distinctive feature of the EMMA platform as it provides built-in automated transcription and translation for all video and text contents. This includes transcription in 7 languages (English, Italian, Spanish, Dutch, French, Portuguese and Estonian) and automatic translation into English, Spanish and Italian. Automatic transcriptions and translations are reviewed by lecturers to reach publishable quality. Most courses have been offered in bilingual (original language plus English) or trilingual form (with Spanish, French or Italian as a third language). Figure 1 shows a unit of a trilingual MOOC in French then translated into English and Italian. A translation button allows to switch between languages.

The UPV media repository

The UPV media repository is a service for the creation, storage, management and dissemination of video lectures, called poliMedias. poliMedias provide a concise overview of a given topic and have an average duration of ten minutes (Turró et al., 2009). Figure 2 shows an example with subtitles in Spanish and English. Table 1 shows basic statistics on poliMedias by their most common languages. poliMedia subtitles can be reviewed anonymously, though editions must be approved by the lecturer before publication.

<table>
<thead>
<tr>
<th>Language</th>
<th>Videos</th>
<th>Hours</th>
<th>Lecturers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish</td>
<td>15013</td>
<td>2709</td>
<td>1572</td>
</tr>
<tr>
<td>English</td>
<td>1221</td>
<td>173</td>
<td>203</td>
</tr>
<tr>
<td>Catalan</td>
<td>434</td>
<td>52</td>
<td>80</td>
</tr>
</tbody>
</table>

Table 1. Number of poliMedia hours of video per language

Systems, tools and integration components

Two main open-source tools have been used to develop ASR and MT systems: the transLectures-UPV Toolkit (TLK) (del-Agua et al., 2014) and the Moses toolkit (Koehn et al., 2007). These systems have been adapted by applying the techniques described in Martínez-Villaronga et al. (2013) and Axelrod et al. (2011). Then, these systems have been integrated into the case studies using the transLectures-UPV Platform (TLP) (see mlp.upv.es/tlp).

Figure 3 shows two examples of use of the multilingual TLP player/editor. The first example (top) is an editor for transcriptions. The video and its segmentation are displayed to the left, while transcriptions are shown to the right. The second example (bottom) is an editor for transcriptions and translations. It is analogous to the first example, but with translations also available to the right. Also, the TLP player/editor can be used to review text documents.
The systems developed by the MLLP research group for the EMMA platform and the UPV media repository can be freely tried through the Transcription and Translation Platform (TTP) (see ttp.mllp.upv.es).

Transcription and translation quality

In this section, we assess the quality of automatic transcriptions and translations generated by the MLLP’s ASR/MT systems for videos originally in 5 languages drawn from the UPV media repository and the EMMA platform. Additionally, a comparative evaluation of transcription and translation quality with mainstream providers of ASR/MT technology, i.e., YouTube and Google Translate, is also presented.

Transcription quality

Transcription quality was measured with the widely accepted Word Error Rate (WER) criterion (Hunt, 1990). Formally, the WER is the normalized minimum number of elementary word editing operations required to transform an automatic transcription into its corrected (reviewed) version. Three elementary word editing operations are considered: insertions, deletions and substitutions. Normalization is computed with respect to the number of words in the reviewed transcription, and often expressed as a percentage. For example, if a lecturer has to apply 30 elementary editing operations to an automatic transcription so as to obtain a reviewed version with a length of 200 words, then the WER will be 15%. In this regard, it must be noted that expecting to achieve error-free transcriptions is unrealistic, even if they are manually produced. On the contrary, it is more realistic to expect a WER of about 10% from commercial, manual transcription services (Hazen, 2006). From a practical point of view, automatic transcriptions of WER equal or less than 25% convey enough correct information to be useful (Munteanu et al., 2006), and professional stenographers prefer them to manually transcribing from scratch (Akita et al., 2009).

Table 2 shows the number of videos, duration (in hours) and WER (± standard deviation) for each transcribed language. Spanish- and English-language videos come from the UPV media repository, while Italian-, Dutch- and French-language videos were included in MOOCs delivered on the EMMA platform. There are a significant number of Spanish-language videos, since more than 90% of the videos in the UPV media repository are in Spanish.

The average duration of videos for all languages except for Dutch is less than 10 minutes. Dutch videos last more than 35 minutes on average and the format of the video presentation is different from that in the other languages. Dutch videos are interviews with usually two speakers sitting around a table, while in the other videos a single speaker stands in front of the camera.
Table 2. Videos, duration (hrs.) and WER (± std. dev.) per language

<table>
<thead>
<tr>
<th>Language</th>
<th>Videos</th>
<th>Hours</th>
<th>WER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish</td>
<td>207</td>
<td>24.7</td>
<td>18.4 ± 6.4</td>
</tr>
<tr>
<td>Italian</td>
<td>13</td>
<td>1.2</td>
<td>25.7 ± 6.4</td>
</tr>
<tr>
<td>English</td>
<td>25</td>
<td>3.5</td>
<td>21.9 ± 8.5</td>
</tr>
<tr>
<td>Dutch</td>
<td>11</td>
<td>6.9</td>
<td>29.4 ± 9.2</td>
</tr>
<tr>
<td>French</td>
<td>21</td>
<td>2.1</td>
<td>23.2 ± 8.3</td>
</tr>
</tbody>
</table>

From the results in Table 2, we can observe that the quality of Spanish transcriptions is the highest, followed by English and French, all three being below 25%. Italian is just above 25% of WER and Dutch has the highest WER figure, but still below 30% of WER. In the case of Dutch, we believe that the higher WER figure is explained by the presence of more than one speaker in the videos, which harms the acoustic adaptation to the speaker, not being so effective as in the rest of the videos in which a single speaker appears.

Translation quality

As with transcription, translation quality is often measured with an error criterion: the so-called Translation Edit Rate (TER) (Snover et al., 2006). This criterion is computed in the same way as the WER, which is, as a normalized percentage of the minimum number of elementary word editing operations required to transform an automatic output (translation) into its reviewed version. The only significant difference is that, apart from insertions, deletions and substitutions, shifts are also allowed. Also as with the WER, it must be noted that achieving error-free translations, either automatic or manual, is unrealistic. Additionally, in the case of MT it is generally accepted that source sentences can be manually translated in many different yet correct ways, and thus a correct translation for a certain reviewer might not be the preferred (correct) translation for another one. As the TER is computed from only one correct reference, it is considered a pessimistic criterion. From a practical point of view, automatic translations with TER figures below 50% are worth post-editing, instead of translating from scratch (Specia et al., 2009; Specia, 2011).

Table 3 shows the number of videos, duration (in hours) and TER (± standard deviation) for each translation pair. All videos were automatically translated and then reviewed. The Spanish-language videos are part of the UPV media repository and were reviewed by lecturers. The English→Spanish videos are from two EMMA MOOCs originally in Italian, then translated into English, and now for this work translated into Spanish. Analogously, the English→Italian videos are from two EMMA MOOCs originally in Spanish, then translated into English, and finally translated into Italian. In this evaluation set there are four MOOCs available in three languages (Italian, English and Spanish). Finally, the Dutch- and French-language videos are also from EMMA MOOCs translated into English.

Table 3. Videos, duration (hrs.) and TER (± std. dev.) per translation pair

<table>
<thead>
<tr>
<th>Translation pair</th>
<th>Videos</th>
<th>Hours</th>
<th>TER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish → English</td>
<td>101</td>
<td>10.8</td>
<td>33.2 ± 14.4</td>
</tr>
<tr>
<td>English → Spanish</td>
<td>29</td>
<td>2.5</td>
<td>27.0 ± 19.9</td>
</tr>
<tr>
<td>Italian → English</td>
<td>14</td>
<td>1.6</td>
<td>37.5 ± 8.2</td>
</tr>
<tr>
<td>English → Italian</td>
<td>121</td>
<td>6.5</td>
<td>33.8 ± 8.0</td>
</tr>
<tr>
<td>Dutch → English</td>
<td>5</td>
<td>3.5</td>
<td>30.7 ± 13.4</td>
</tr>
<tr>
<td>French → English</td>
<td>8</td>
<td>0.9</td>
<td>58.9 ± 5.2</td>
</tr>
</tbody>
</table>

From the results in Table 3, it is clear that, apart from the French→English pair, the translation quality is good enough to be worth post-editing (below 50% TER). The translation quality of the French→English pair was lower than expected. This phenomenon is due mainly to two reasons. First, the reviewers used a two-pass review process when generating the final translations that makes them differ significantly from those that would be obtained in a single pass, as was the case with the other translation pairs. Second, we believe that the MT system providing the automatic English translations from French did not properly adapt to the specific domain of the French courses.

Comparison with mainstream providers

One of the questions that arises is how the adapted systems deployed in this work compare to systems from mainstream providers and, in particular, to the state-of-the-art YouTube automatic captioning and Google
Translate systems. For this purpose, a different evaluation set was defined with videos from MOOCs offered in the EMMA platform. Table 4 shows, for each transcribed language, the number of videos included in this evaluation set, their duration, and the WER achieved by the MLLP’s TTP and YouTube’s automatic captioning.

Table 4. Videos, duration (hrs.), and TTP and YouTube WER per language

<table>
<thead>
<tr>
<th>Language</th>
<th>Videos</th>
<th>Hours</th>
<th>TTP</th>
<th>YouTube</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish</td>
<td>23</td>
<td>3.5</td>
<td>14.8</td>
<td>22.5</td>
</tr>
<tr>
<td>Italian</td>
<td>3</td>
<td>4.0</td>
<td>17.1</td>
<td>31.6</td>
</tr>
<tr>
<td>English</td>
<td>9</td>
<td>0.4</td>
<td>39.2</td>
<td>65.9</td>
</tr>
<tr>
<td>Dutch</td>
<td>2</td>
<td>1.1</td>
<td>24.5</td>
<td>41.1</td>
</tr>
<tr>
<td>French</td>
<td>18</td>
<td>2.3</td>
<td>20.6</td>
<td>32.0</td>
</tr>
</tbody>
</table>

From the results in Table 4, we can conclude that YouTube’s WER is higher than that of TTP’s systems for all languages, and more precisely, the relative WER increase over TTP’s is nearly 70% on average. The main reason behind these results is the fact that YouTube uses general-purpose ASR systems, while the ASR systems integrated into TTP are automatically adapted to the task as described in Section 3. The English ASR system obtained a surprisingly high WER compared with the one reported in Table 2 based on the same technology. An error analysis on the English videos studied in this work led to the conclusion that the accent of the only speaker in these videos was especially difficult to understand.

The evaluation set used to compare transcriptions was enlarged for the purpose of comparing translations. Table 5 shows, for each translation pair, the number of videos included in the translation evaluation set, their duration, and the TER obtained with the MLLP’s TTP and Google Translate. These videos were previously transcribed in order to be translated. In the case of English into Spanish and French into English, the same English- and French-language videos transcribed in Table 4 were then translated.

Table 5. Videos, duration (hrs.), and TTP and Google TER per translation pair

<table>
<thead>
<tr>
<th>Translation pair</th>
<th>Videos</th>
<th>Hours</th>
<th>TTP</th>
<th>Google</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish → English</td>
<td>250</td>
<td>13.9</td>
<td>33.9</td>
<td>44.3</td>
</tr>
<tr>
<td>English → Spanish</td>
<td>9</td>
<td>0.4</td>
<td>35.8</td>
<td>42.4</td>
</tr>
<tr>
<td>Italian → English</td>
<td>11</td>
<td>1.1</td>
<td>33.4</td>
<td>39.2</td>
</tr>
<tr>
<td>English → Italian</td>
<td>81</td>
<td>5.4</td>
<td>39.7</td>
<td>43.3</td>
</tr>
<tr>
<td>Dutch → English</td>
<td>2</td>
<td>1.2</td>
<td>42.5</td>
<td>45.0</td>
</tr>
<tr>
<td>French → English</td>
<td>18</td>
<td>2.3</td>
<td>52.8</td>
<td>52.6</td>
</tr>
</tbody>
</table>

A general conclusion that can be drawn from Table 5 is that Google Translate’s MT systems produce a higher translation error than TTP’s MT systems, except for French into English, where both systems show a similar performance. On average, the TER figures achieved by Google Translate are higher than those of TTP by 14% relative. Again, as opposed to the general-purpose MT systems provided by Google Translate, TTP systems are adapted to the domain of the video that is being translated, and thus more accurate results are obtained.

**Reviewing time**

The time required for reviewers (e.g., lecturers) to post-edit automatic video transcriptions and translations is measured in terms of Real Time Factor (RTF) (Valor-Miró et al., 2015). This measure is the video duration-normalized time required for the reviewer to post-edit the whole video transcription (or translation). For instance, if a video lasts 6 minutes and its review takes one hour (60 minutes), then the RTF will be 10.

In general, manual annotation of speech ranges from 10 RTF, in the case of orthographic transcription (Reidsma et al., 2005), to 50 RTF, in which a detailed 4-level speech annotation is performed (Barras et al., 2001). Expert transcriptionists can achieve as low an RTF as 6 (Williams et al., 2011), but this is not the usual profile for lecturers. In our previous work, (Valor-Miró et al., 2015), the RTF for manual (orthographic) transcription attained by lecturers was 10.1 ± 1.8, which matches the figures reported in (Reidsma et al., 2005). For this reason, we take 10 RTF as a reference review time for transcription.

Regarding the RTF for translation, in contrast to transcription, it is more difficult to establish a single reference RTF, except for the rule of thumb of 2500 words per day of work, since translation is a more complex task requiring a greater cognitive effort and involving different factors such as source and target languages, degree of expertise and experience of the translator, vocabulary specificity, software tools, etc. Having in mind this
limitation, specialist translators achieve fully-manual translating rates ranging from 400 to almost 1000 words per hour (Plitt & Masselot, 2010). Taking these figures into the UPV media repository in which speakers utter 150 words per minute on average, a specialist translator would be translating at 22.5 RTF in the worst case. In the transLectures project (Turró et al., 2016), seven hours of videos drawn from the UPV media repository were translated ex novo from Spanish into English by two professional translators achieving an average RTF of 34.1 ± 11.4 RTF. For the sake of comparison and taking into account the profile of the translators in this case (lecturers), hereinafter we consider the RTF of manual translation to be 30 RTF.

Transcription reviewing time

Table 6 shows, for each transcribed language, the average WER (copied from Table 2) and RTF (± std. dev.), and regression models to predict RTF as a function of WER. Three regression models were tried: linear, square root and logarithm. In the case of Spanish, detailed information is provided in Table 6 on the adjustment of these three regression models. Also, Figure 4 shows a scatter plot of RTF (y axis) versus WER (x axis) for each Spanish-language video (plotted point) and each adjusted regression model. For the rest of the transcribed languages, only the details on the adjustment of the logarithmic model are given in Table 6 for brevity.

Table 6. Average WER and RTF (± std. dev.), and regression models per language

<table>
<thead>
<tr>
<th>Language</th>
<th>WER</th>
<th>RTF ± std. dev.</th>
<th>Model</th>
<th>$R^2$</th>
<th>$\beta$</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish</td>
<td>18.4</td>
<td>3.3 ± 1.2</td>
<td>WER</td>
<td>0.87</td>
<td>0.17</td>
<td>$&lt; 10^{-15}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\sqrt{\text{WER}}$</td>
<td>0.90</td>
<td>0.78</td>
<td>$&lt; 10^{-15}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ln WER</td>
<td>0.91</td>
<td>1.17</td>
<td>$&lt; 10^{-15}$</td>
</tr>
<tr>
<td>English</td>
<td>21.9</td>
<td>5.3 ± 1.7</td>
<td>ln WER</td>
<td>0.92</td>
<td>1.76</td>
<td>$&lt; 10^{-14}$</td>
</tr>
<tr>
<td>Italian</td>
<td>25.7</td>
<td>3.9 ± 1.4</td>
<td>ln WER</td>
<td>0.90</td>
<td>1.20</td>
<td>$&lt; 10^{-6}$</td>
</tr>
<tr>
<td>Dutch</td>
<td>29.4</td>
<td>5.8 ± 2.5</td>
<td>ln WER</td>
<td>0.85</td>
<td>1.75</td>
<td>$&lt; 10^{-14}$</td>
</tr>
<tr>
<td>French</td>
<td>23.2</td>
<td>6.7 ± 0.8</td>
<td>ln WER</td>
<td>0.98</td>
<td>2.17</td>
<td>$&lt; 10^{-15}$</td>
</tr>
</tbody>
</table>

Figure 4. RTF vs. WER for Spanish-language videos and prediction models

A first important conclusion from the results on transcription reviewing time is that the availability of automatic transcriptions reduces between one third and two thirds the time devoted to generate video transcriptions. Generally speaking, we may say that the RTF is between 3 and 7 when starting from automatic transcriptions that are worth post-editing, as shown in Table 6. The second important conclusion is that the logarithmic regression model provides a good, statistically significant fit of the observed data, better indeed than the other two models considered. The logarithmic model explains better the fact that users tend to ignore automatic transcriptions when the corresponding WER is too high and prefer retranscribing from scratch to correcting a low-quality automatic transcription. For all languages, the adjustment is statistically significant (Sig. < $10^{-5}$) and an important amount of the variability of the data is explained by the model ($R^2 \geq 0.85$).

On a per-language analysis, Dutch presents higher RTF figures than Spanish, Italian and English. We believe this is explained by the interview format of these videos. Finally, the RTF figure for French is not the one expected from the WER figure reported; indeed, this RTF figure is the highest in this transcription evaluation. The reason behind this RTF figure is the two-pass review process that lecturers carried out in this case. The second pass in the review process requires at least 1 additional RTF, which is the minimum amount of time required to watch the entire video again.
Translation review time

Table 7 shows, for each translation pair, the average TER (copied from Table 3) and RTF (± std. dev.), and regression models to predict RTF as a function of TER. Translation results are provided in Table 7 and Figure 5, in the same way as above for transcription.

<table>
<thead>
<tr>
<th>Translation pair</th>
<th>TER</th>
<th>RTF ± std. dev.</th>
<th>Model</th>
<th>$R^2$</th>
<th>$\beta$</th>
<th>sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish → English</td>
<td>33.2</td>
<td>9.1 ± 4.9</td>
<td>TER</td>
<td>0.75</td>
<td>0.25</td>
<td>$&lt; 10^{-15}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\sqrt{\text{TER}}$</td>
<td>0.80</td>
<td>1.61</td>
<td>$&lt; 10^{-15}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ln TERM</td>
<td>0.80</td>
<td>2.71</td>
<td>$&lt; 10^{-15}$</td>
</tr>
<tr>
<td>English → Spanish</td>
<td>27.0</td>
<td>7.8 ± 4.9</td>
<td>ln TER</td>
<td>0.82</td>
<td>2.67</td>
<td>$&lt; 10^{-11}$</td>
</tr>
<tr>
<td>Italian → English</td>
<td>37.5</td>
<td>11.3 ± 4.2</td>
<td>ln TER</td>
<td>0.89</td>
<td>3.15</td>
<td>$&lt; 10^{-7}$</td>
</tr>
<tr>
<td>English → Italian</td>
<td>33.8</td>
<td>9.6 ± 5.3</td>
<td>ln TER</td>
<td>0.77</td>
<td>2.76</td>
<td>$&lt; 10^{-15}$</td>
</tr>
<tr>
<td>Dutch → English</td>
<td>30.7</td>
<td>9.5 ± 3.9</td>
<td>ln TER</td>
<td>0.91</td>
<td>2.89</td>
<td>$&lt; 10^{-2}$</td>
</tr>
<tr>
<td>French → English</td>
<td>58.9</td>
<td>23.2 ± 8.0</td>
<td>ln TER</td>
<td>0.90</td>
<td>5.67</td>
<td>$&lt; 10^{-4}$</td>
</tr>
</tbody>
</table>

Figure 5. RTF vs. TER for Spanish into English videos; and prediction models

Similarly to transcription, the first important result is that, except for French→English, the review time is reduced to approximately one third when the quality of the automatic translations is worth post-editing, as shown in Table 7. The second result is that the logarithmic regression model is among the best explaining the observed data. Again, the logarithmic model fits better the cases with high values of TER, bounding the corresponding RTF, since reviewers ignore those automatic translations containing too many errors and prefer to generate the translation from scratch. The amount of the variability of the data explained by the model ($R^2$ values) is not as high as in the review of transcriptions, which is reflected in Figure 5 as a greater dispersion of the data points. The reason behind this behaviour is the higher complexity of the translation task (compared to transcription), which involves a significant cognitive load.

In a per-translation-pair analysis, the review of Spanish translations from English transcriptions is similar to the translation in the opposite translation pair, but the RTF figure is even lower for the latter. This fact correlates with the Italian into English and English into Italian translation pairs, since most of the reviewers involved are non-native English speakers, and it is easier for them to translate into their mother tongue. The figures for the Dutch into English translation review are very much in line with the previous translation pairs, considering that the quality of the automatic translations was among the best. Finally, the translation of French courses was surprisingly cumbersome, taking far more time than the other translation pairs. This phenomenon is due mainly to two reasons. First, as mentioned above, the MT system that generated the automatic English translations from French did not properly adapt to the domain of the courses; and second, reviewers employed a two-pass review process that was more costly than the conventional one-pass review process used in the rest of translation pairs.

Review time across languages

In the previous sections we have found that, for each language involved, a logarithmic regression model can be adjusted to accurately predict RTF from transcription WER; and we have reached a similar conclusion in translation (i.e., to predict RTF from TER) for each translation pair assessed. Therefore, it is worth asking whether a single logarithmic regression model could suffice to accurately predict RTF from WER (TER) across
all languages (translation pairs) under study. This is considered in Figure 6. The scatter plot at the top shows RTF versus WER, for all languages involved (plotted points), and a single logarithmic regression model fitted to data (videos) pooled across languages. The scatter plot at its bottom is similar, but for TER.

As for predicting RTF from transcription WER, the fitted logarithmic model shown at the top of Figure 6 ($R^2 = 0.87$, $\beta = 1.34$) is statistically significant ($\text{Sig.} < 10^{-15}$). This confirms that the review time depends highly on transcription quality and, to a lesser extent, on the language considered. It is worth noting, however, that most data points (videos) are for Spanish (207 out of 277), and thus results are certainly biased towards this language. In this regard, a closer look at the distribution of data points reveals that they are more or less clustered by language. This was not unexpected since, after all, there are language- and MOOC-dependent factors (e.g., topic, reviewers and review quality requirements) that certainly have some effect on the RTF but fall out of the scope of this work. In any case, the statistical significance of the fit suffices to support the idea that RTF mainly depends on WER, irrespective of the transcription language. For example, and to be more precise, taking a couple of reference points on the logarithmic curve we can infer that a one-hour video transcription of 10 WER points will take 3 hours to be reviewed, and a video of the same duration with 20 WER points of transcription error will require almost 4 hours. This is significantly less time than the 10 RTF for transcribing from scratch.

![Figure 6. RTF vs. WER per transcription language (left) and RTF vs. TER per translation pair (right)](image)

As with WER, the fitted logarithmic model shown at the bottom of Figure 6 ($R^2 = 0.78$, $\beta = 2.90$) is statistically significant ($\text{Sig.} < 10^{-15}$) for RTF prediction from TER. As above, then, we can confirm that RTF depends more on the translation quality (TER) than on the language pair considered. In contrast to the above results for WER, however, the distribution of data points does not reveal a clear language pair-dependent clustering structure. Taking into account that data points for Spanish (i.e., Spanish→English) are still dominant (250 out of 371), this adds more evidence to support the validity of the fitted logarithmic model. If, for example, a reviewed one-hour video transcription is automatically translated with about 30 TER points, then we may expect an RTF of around 9, that is, 9 hours for reviewing the translation. This is much less than the 30 hours (30 RTF) we may expect if translation is carried out manually from scratch; in other words, it entails a review time saving of 70% relative.

**Impact on the case studies**

Over the last two years, we have been collecting precise statistics on multilingual data consumption in the two real-life case studies mentioned above: the EMMA platform and the UPV media repository. This data is summarized below in order to better gauge the impact that the availability of video transcriptions and translations has had on both case studies.

**The EMMA platform**

Table 8 shows the number of native and non-native students enrolled in MOOCs offered on the EMMA platform, organized by the original language of the course. It goes without saying that non-native students could only follow these MOOCs thanks to the TLP-based multilingual component in EMMA described above. The last column in Table 8 shows the relative increase in the total number of students over native students due to the enrolment of non-native students.
Table 8. Statistics on student enrolment in MOOCs on the EMMA platform

<table>
<thead>
<tr>
<th>Language</th>
<th>Native students</th>
<th>Non-native students</th>
<th>Relative increase (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish</td>
<td>161</td>
<td>547</td>
<td>340</td>
</tr>
<tr>
<td>French</td>
<td>983</td>
<td>879</td>
<td>89</td>
</tr>
<tr>
<td>Italian</td>
<td>609</td>
<td>259</td>
<td>43</td>
</tr>
<tr>
<td>Dutch</td>
<td>501</td>
<td>104</td>
<td>21</td>
</tr>
<tr>
<td>English</td>
<td>351</td>
<td>27</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>2605</td>
<td>1816</td>
<td>70</td>
</tr>
</tbody>
</table>

Note that the results in Table 8 are given in decreasing order respect to the relative increment of non-native students. The best results were obtained by MOOCs originally in Spanish and followed by 161 Spanish-speaking students. As these courses were also delivered in English and Italian, 547 non-Spanish-speaking students enrolled in the courses, increasing the total number of students by 340% with respect to the Spanish-speaking students. MOOCs in French almost doubled their number of students by offering these courses also in English. MOOCs in Italian and Dutch translated into English also experienced a relative increase with the non-native students enrolled of approximately 40% and 20%, respectively. Finally, English courses translated into Spanish had a small relative increase in student enrolment, mainly explained by the fact that English is considered a lingua franca and many non-native students are able to follow the course in English, at least students at this level of education. Overall, the translated versions of the MOOCs facilitated by the TLP in the EMMA platform attracted students that are non-native in the original language of the courses, increasing the total student enrolment by a notable 70%.

Indeed, according to exit questionnaires filled in by almost 1500 students enrolled in EMMA courses, 75% of them appreciated multilinguality as a feature of this platform and 70% found multilingual subtitles useful (Ferrari et al., 2016a). Taking into account only those approximately 200 students that replied to mini-questionnaires embedded in 17 running MOOCs, 31% of them always used the translation functionality, that is, the MOOC was originally in a different language from their mother tongue; and 29% of them sometimes used the translation functionality. Indeed, at least 90% of the students using always or sometimes the translation functionality agreed that this functionality enhances the overall value of the EMMA platform and makes EMMA a truly European experience (Ferrari et al., 2016b).

The UPV media repository

Table 9 shows the number of poliMedia videos and subtitle views (in thousands) per language and in total from June 2015, when view logs were activated, to May 2016.

<table>
<thead>
<tr>
<th>Video language</th>
<th>Video views</th>
<th>Subtitle views</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Spanish</td>
<td>English</td>
</tr>
<tr>
<td>Spanish</td>
<td>629</td>
<td>6.9</td>
<td>1.1</td>
</tr>
<tr>
<td>English</td>
<td>63</td>
<td>1.3</td>
<td>0.5</td>
</tr>
<tr>
<td>Total</td>
<td>692</td>
<td>8.2</td>
<td>1.6</td>
</tr>
</tbody>
</table>

The main conclusion that can be drawn from Table 9 is that, on average, subtitles were turned on in 1.4% of video views. It is worth noting, however, that a 1.4% of a large number of video views (i.e., almost 700K over the last year) is a significant number of users turning subtitles on (i.e., almost 10K over the last year). Indeed, in relative terms, it is interesting to observe that 2.5% of the English-language videos had their subtitles activated, in contrast to Spanish-language videos which did in 1.3% of the views. This result does not come as a surprise since most UPV students are native Spanish speakers with English as a foreign language. Finally, Spanish subtitles were predominant when subtitles were activated, being chosen in 86% and 71% of the cases for Spanish- and English-language videos, respectively.

Apart from the accessibility benefits for hearing-impaired and foreign students, the availability of transcriptions has allowed for the indexing and subsequent search for specific words in this large video lecture repository. Indeed, this search tool at the UPV media repository allows students to find the specific video clip in which a word is uttered by the lecturer. Thus, students can discard video clips that are not of their interest to focus on those ones in which a specific concept is explained, saving a significant amount of time. Subtitles also provide support for students in their arduous note-taking tasks.
Conclusions

In this work, we have reported a large part of the experience we have gained from producing low-cost multilingual video subtitles of publishable quality for MOOCs and OER. Apart from describing the systems, tools and integration components employed for such purpose, a comprehensive evaluation of the results achieved has been provided from three viewpoints: the quality of video transcriptions and translations automatically generated from task-adapted ASR/MT systems, the time required to review them, and the impact multilingual subtitles have had on a MOOC platform and a large video lecture repository.

The quality of automatic transcriptions and translations has been proved to be in most cases below 25% of WER and 50% of TER, respectively. This means that it is worth post-editing them to achieve publishable subtitles instead of generating them ex novo. Indeed, the output of the adapted ASR/MT systems has been positively compared to state-of-the-art automatic transcription and translation tools provided by mainstream providers. More precisely, these systems are on average 38% and 17% better than YouTube’s automatic captioning and Google Translate, respectively.

Regarding the review process, we have showed that a lecturer can save between 30% and 70% of the time devoted to review transcriptions, and between 25% and 75% of the translation review time, with respect to performing these tasks from scratch. In addition, a multilingual linear regression model has been proposed to infer the review time (RTF) as a function of WER in the case of transcription, and in terms of TER for translation.

The availability of multilingual video subtitles has been shown to have a great impact in our case studies. On the one hand, in the EMMA platform, the translation of MOOCs into a second, or even a third language has significantly increased course visibility boosting student enrolment by 70% relative. On the other hand, multilingual subtitles at the UPV media repository have not only improved accessibility to the video lectures for hearing-impaired and non-native-speaking students, but also have allowed the development of added-value functionalities such as indexing and search capabilities, and obviously translated subtitles.

Acknowledgements

The research leading to these results has received funding from the European Union’s Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 287755 (transLectures) and from the EU’s ICT Policy Support Programme as part of the Competitiveness and Innovation Framework Programme under grant agreement no. 621030 (EMMA). Additionally, it is supported by the Spanish research project TIN2015-68326-R (MINECO/FEDER).

References


A Structural Model for Students’ Adoption of Learning Management Systems: An Empirical Investigation in the Higher Education Context

Duygu Fındık-Coşkunçağ, Nurcan Alkış, and Sevgi Özkan-Yıldırım

1Faculty of Economics and Administrative Sciences, Atatürk University // 2Technology and Knowledge Management, Başkent University // 3Information Systems, Middle East Technical University

dygfndk@gmail.com // nalkis@baskent.edu.tr // sevgiozk@metu.edu.tr

*Corresponding author

(Submitted October 12, 2016; Revised December 01, 2016; Accepted January 31, 2017)

ABSTRACT
With the recent advances in information technologies, Learning Management Systems have taken on a significant role in providing educational resources. The successful use of these systems in higher education is important for the implementation, management and continuous improvement of e-learning services to increase the quality of learning. This study aimed to identify the factors affecting higher education students’ behavioral intention towards Learning Management Systems. A research model was proposed based on the belief factors of the technology acceptance model; namely perceived usefulness, perceived ease-of-use and external factors including self-efficacy, enjoyment, subjective norm, satisfaction, and interactivity and control. Then, a self-reported questionnaire was distributed online. A total of 470 higher education students participated in the survey. The proposed structural model was assessed and validated using structural equation modeling, in particular the partial least square method. The predictors of behavioral intention were identified as perceived usefulness, perceived ease-of-use, enjoyment, subjective norm, satisfaction, and interactivity and control with the validated structural model. The relationships between the influencing factors provided an insight about the students’ behavioral intention towards the use of Learning Management Systems. It is expected that the academicians and practitioners will benefit from the design and findings of the current study in their future research.

Keywords
Learning management systems, Student adoption, Technology acceptance model, Structural equation modeling, Partial least square

Introduction
With the advances in information and communication technologies, educational activities are now more dependent on the internet and online applications. These new developments have resulted in the emergence of a new concept, e-learning. E-learning refers to “technology-based learning in which learning materials are delivered electronically to remote learners via a computer network” (Zhang, Zhao, Zhou, & Nunamaker Jr, 2004, p. 76). Several applications are used to support e-learning activities; such as course, learning and student management systems, accounting systems, content creation tools and course websites (Paulsen, 2003). A Learning Management System (LMS) is one of the widely used applications in higher education institutions to support course activities in the digital environment. The effective implementation of this tool is important to improve the quality of learning, access to education and training, provide cost-effectiveness and reduce the cost of education (Bates, 1997). However, contrary to expectations, the implementation of this system may be problematic, often resulting in failure (Bhuiasiri, Xaymoungkhoun, Zhuo, Rho, & Ciganek, 2012). Therefore, the problems and challenges involved in the adoption and implementation of LMS should be investigated.

The effective use of LMS in the education field mainly depends on certain factors related to the behavioral attitudes of instructors and students, university support and applied information technologies (Davis, Bagozzi, & Warshaw, 1989; Webster & Hackley, 1997). In particular, the users of these systems may have a different point of view towards technology adoption and acceptance; therefore, this is important to consider when evaluating technology-mediated online learning systems (Dillon & Gunawardena, 1995). In the education field, instructors and students are the end users of LMS; thus, they play a major role in the successful implementation of this system. Since students are the main target group to benefit from LMS, their adoption of this system is important, particularly in higher education. In this context, this study aimed to identify the key factors affecting students’ behavioral intention towards the use of LMS, namely NET-ClassR, in higher education by taking Technology Acceptance Model (TAM) as the theoretical basis. NET-ClassR was designed to meet the e-learning needs and manage courses without the requirement of extensive technical knowledge. NET-ClassR has three main users, the instructors, students and the administrator. It provides separate functions and graphical user interfaces for each type of user. The users can follow and manage web-based asynchronous courses using a web interface.
In the literature, there is a considerable amount of research on the acceptance of e-learning applications by students. In particular, recent studies have addressed the adoption of synchronous and asynchronous web-based technologies (Lee, Yoon, & Lee, 2009), e-learning systems (Lee, 2010; Pituch & Lee, 2006), web course tools (WebCT) (Sánchez-Franco, 2010), web-based streaming media (Liu, Liao, & Pratt, 2009), web-based learning systems (Lee, 2008; Saadé & Bahli, 2005), virtual learning environments (Van Raaij & Schepers, 2008), web-based educational tools (Ngai, Poon, & Chan, 2007), e-learning courses (Park, 2009), web-based class management systems (Yi & Hwang, 2003), discussion forum (Aucamp & Swart, 2015) and LMS (Murshitha & Wickramarachchi, 2016). However, research on the acceptance of LMS in blended learning environments remains relatively limited; therefore, conducting studies on acceptance of LMS to support traditional learning would contribute to the literature. The current study was based on the research question, “what are the factors influencing students’ acceptance of LMS?” This paper presents the behavioral intention of students in relation to LMS, NET-ClassR, via a new research model using the new LMS-TAM.

In the literature, different theoretical frameworks and research models have been developed and used to evaluate the individuals’ adoption or rejection of new technologies. Therefore, determining the influencing factors of the proposed research model is a challenging task. In addition to the in-depth systematic review of literature (Alkış, Fındık-Coşkunçay, & Özkan-Yıldırım, 2014), experts from academia were employed to assist in the development of the model. The constructs of the model were identified as: Perceived Usefulness (PU), Perceived Ease-of-Use (PEOU), Behavioral Intention (BI), Self-Efficacy (SE), Enjoyment (ENJ), Subjective Norm (SN), Satisfaction (STS), and Interactivity and Control (IC).

**Theoretical background**

**Concept of e-learning and its advantages**

In the information age, e-learning, also referred to as web-based learning, is one of the most popular learning environments (Liaw, Huang, & Chen, 2007). E-learning systems help use time and space efficiently; however, their success depends on end users’ acceptance and use of these systems (Van Raaij & Schepers, 2008). In order to support e-learning activities, several technology-based pedagogical tools have been developed; such as web course tools, web course homepage system, blackboard learning system and system for multimedia integrated learning (Ngai et al., 2007).

E-learning provides many benefits including an increased accessibility to information, better content delivery, personalized instruction, content standardization, accountability, on-demand availability, self-pacing, interactivity, confidence, and increased convenience (Bhuasiri et al., 2012). As benefits from e-learning systems depend on users’ adoption and continued use (Tai, Zhang, Chang, Chen, & Chen, 2012), users’ adoption of this technology needs to be examined with the help of behavioral intention theories. Therefore, it is important to understand the predictive factors of the students’ behavioral intention to use e-learning systems.

**Technology acceptance model**

Although information technology has grown dramatically, there is a considerably high level of resistance in end users to using e-learning applications. Many researchers have studied the behavioral intention of end users towards new technologies to reveal the dimensions affecting adoption or rejection of these technologies. These studies use TAM (Davis et al., 1989) as theoretical base since it is the most effective model in providing an understanding and predicting the acceptance of information technology.

TAM was developed by Davis in 1986 as an adapted version of the Theory of Reasoned Action (TRA) for the technology domain. TAM proposes that technology use is determined by behavioral intention, which is determined by perceived usefulness, perceived ease-of-use and attitude (Davis et al., 1989). This model is theoretically justified and provides an insight into end-user behavior across a broad range of computing technologies (Lee, Cheung, & Chen, 2005).
Methodology

Research model

In this study, a structural research model LMS-TAM was proposed to predict students’ adoption of LMS (Figure 1) by using TAM as the theoretical framework. The main aim of this model is to identify actual behavior with behavioral intention. To reach the actual behavior, it is important to identify behavioral intention and its direct predictors. Therefore, in the proposed model, attitude was excluded since it is aimed to identify the direct predictors of behavioral intention from external factors, differently from TAM. Additional factors and hypotheses were formed following the recommendations of Ma, Andersson and Streith (2005). The researchers emphasized that TAM included only two key explanatory factors that are PU and PEOU; for this reason, it is insufficient to fully understand the relations between information systems and users acceptance behavior. Therefore, additional factors and their relations were considered to increase predictive power of the model. After a systematic review of the literature on e-learning (Alkış et al., 2014), a number of theories and behavioral constructs were selected, examined and categorized by three experts with experience in the subject area. Then, the relationships between these constructs were explored.

For the development of the model, card-sorting and group discussion methods were used and eight constructs (BI, PU, PEOU, STS, ENJ, SN, SE and IC) were identified. The definitions of the selected constructs are given in Table 1. The reliability and validity of the constructs were assessed by a pilot study. Then, hypotheses were proposed related to the relationships between these constructs in accordance with the findings from the literature and experts’ opinions.

In the literature, PU, PEOU and BI are the major determinants of the acceptance of e-learning systems based on TAM. Thus, the following three hypotheses were proposed to assess the effects of these constructs on the acceptance of LMS:

H1: PU directly and positively affects STS.
H2: PU directly and positively affects BI.
H3: PEOU directly and positively affects PU.

In the proposed model, TAM was extended to include SE, ENJ, SN, STS, and IC constructs. A total of 10 hypotheses were formulated to examine the effect of each construct on LMS use:

H4: STS directly and positively affects BI.
H5: ENJ directly and positively affects PU.
H6: ENJ directly and positively affects PEOU.
H7: ENJ directly and positively affects STS.
H8: ENJ directly and positively affects BI.
H9: SN directly and positively affects PU.
H10: SN directly and positively affects BI.
H11: Self-efficacy directly and positively affects PEOU.
H12: Self-efficacy directly and positively affects BI.
H13: IC directly and positively affects PU.

Instrument development

A comprehensive survey was implemented to collect data. The survey instrument consisted of two parts. The first part contained eight questions for demographic data including gender, age, department, education level, experience and competency regarding computer use, familiarity with LMSs, and preferred learning style. The second part was based on a five-point Likert scale (1- “strongly disagree” to 5 “strongly agree”) comprising 44 items that measured the factors of the proposed research model. The items in the second part of the survey were adopted from the scales used in the literature (Table 2). The content validity of the instrument was assessed by an expert panel.
Figure 1. Proposed Research Model (LMS-TAM)

Table 1. Model constructs, definitions and references

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Definition</th>
<th>Prior theories</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived Usefulness</td>
<td>“The degree to which a person believes that using a particular system would enhance his or her job performance”</td>
<td>TAM</td>
<td>Davis, 1989</td>
</tr>
<tr>
<td>Perceived Ease of Use</td>
<td>“The degree to which a person believes that using a particular system would be free of effort”</td>
<td>TAM</td>
<td>Davis, 1989</td>
</tr>
<tr>
<td>Behavioral Intention</td>
<td>“An individual’s performing a conscious act, such as deciding to accept (or use) a technology”</td>
<td>TAM</td>
<td>Davis, 1989</td>
</tr>
<tr>
<td>Self-Efficacy</td>
<td>“The belief an individual has in his/her ability to successfully perform a certain behavior”</td>
<td>Social Cognitive Theory</td>
<td>Bandura, 1986</td>
</tr>
<tr>
<td>Enjoyment</td>
<td>“The extent to which the activity of using a specific system is perceived to be enjoyable in its own right, aside from any performance consequences resulting from system use”</td>
<td>Self-determination theory - Intrinsic Motivation</td>
<td>Venkatesh, 2000</td>
</tr>
<tr>
<td>Subjective Norm</td>
<td>“The social pressure from the social environment on the users to use a system”</td>
<td>TRA</td>
<td>Ajzen, 1991</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>The extent to which a user is pleased or contented with the information system.</td>
<td>D&amp;M Information Systems Success Model</td>
<td>Delone &amp; McLean, 2003</td>
</tr>
<tr>
<td>Interactivity and Control</td>
<td>The system characteristics by which user could interact with each other and control the form and content of a mediated environment.</td>
<td>No prior theories</td>
<td>Martínez-Torres et al., 2008; Steuer, 1992</td>
</tr>
</tbody>
</table>

Table 2. The constructs, items and sources from which the items were adopted

<table>
<thead>
<tr>
<th>Construct</th>
<th>Code</th>
<th>Item</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>PU</td>
<td>Item1</td>
<td>Using NET-ClassR improves my performance in courses.</td>
<td>Davis, 1989</td>
</tr>
<tr>
<td></td>
<td>Item2</td>
<td>I think it is useful to support courses with NET-ClassR.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Item3</td>
<td>NET-ClassR helps me effectively perform my learning activities.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Item4</td>
<td>NET-ClassR is useful to follow course activities online.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Item5</td>
<td>Through the internet connection, NET-ClassR provides several advantages in terms of solving time- and location-related problems.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Item6</td>
<td>NET-ClassR improves my success in courses.</td>
<td>Davis, 1989</td>
</tr>
<tr>
<td>PEOU</td>
<td>Item7</td>
<td>NET-ClassR is easy to use.</td>
<td></td>
</tr>
<tr>
<td>Item</td>
<td>Statement</td>
<td>Source</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-------------------------------</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>It is easy for me to learn to operate the NET-ClassR system.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>It is easy for me to become skillful at using the NET-ClassR system.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Interacting with the e-learning system does not require a lot of mental effort.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>I found it easy to get the e-learning system to implement what I wanted.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>In future, the courses should be supported with NET-ClassR.</td>
<td>Davis, 1989</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>If I have access to NET-ClassR, I intend to use it.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>If courses are supported with NET-ClassR, I intend to use it frequently.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>I think the instructors should support the use of NET-ClassR.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>I think the instructors should continue to use NET-ClassR.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>I am confident of using NET-ClassR even if there is no one around to show me how to do it.</td>
<td>Compeau &amp; Higgins, 1995</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>I am confident of using NET-ClassR even if I do not have an online manual for reference.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>I am confident of using NET-ClassR even if I have never used such a system before.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>I am confident of using NET-ClassR even if I do not watch someone use it before trying it myself.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>SE5: I could complete the learning activities using NET-ClassR even if I could not call anyone for help when I got stuck.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>I find it enjoyable to use NET-ClassR.</td>
<td>Lee et al., 2005</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>I find it interesting to use NET-ClassR.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>I find the interface of NET-ClassR enjoyable.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>NET-ClassR is a fun activity.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>The use of NetClassR arouses my curiosity.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>My instructors’ opinion about the use of NET-ClassR is important for meNET-ClassR.</td>
<td>Taylor &amp; Todd, 1995a; 1995b</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>My instructors think that we should use NET-ClassR.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>My classmates think that I should use NET-ClassR.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>My classmates’ opinion has an effect on my decision to use NET-ClassR.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>The course assistants think that I should use NET-ClassR.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>The assistants’ opinion has an effect on my decision to use NET-ClassR.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>The school management encourages students to use NET-ClassR.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>I am satisfied with the performance of NET-ClassR in helping me follow the courses.</td>
<td>Bhattacherjee, 2001a; 2001b</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>NET-ClassR is a satisfactory system to perform course activities.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>I am satisfied with the courses conducted with the support of NET-ClassR.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>The tools in the NET-ClassR are satisfactory to follow courses.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>In general, supporting courses with NET-ClassR is satisfying.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>NET-ClassR is a satisfactory system to encourage interactive learning.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>NET-ClassR enables interactive communication between the instructor and students.</td>
<td>Martinez-Torres et al., 2008</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>NET-ClassR facilitates interactive communication between students.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>Communication tools in NET-ClassR (chat, e-mail, and forum) are effective in facilitating interactivity between the users.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>NET-ClassR provides an opportunity to control communication between instructors whenever students require.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>NET-ClassR allows controlling the learning sequence.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Study setting

This study was carried out in the Middle East Technical University (METU), one of the leading universities in Turkey. The participants were METU students, users of NET-ClassR, which was used in METU from 1997 to 2014 as an LMS. The home and master pages of the system are shown in Figure 2 and Figure 3, respectively. By providing several tools, this LMS was used to support both traditional and completely online classes. Initially, this system organized and managed lecture notes and provided platforms for discussion and electronic mail. In addition, it offered the possibility for the evaluation of students through quizzes, assignments and exams. Instructors were able to track the students’ participation in discussions and their access to lecture notes. The system provided statistical data about the students’ achievement. Furthermore, the system was able to back up the entire course information including forums, discussions, assignments, grades and lecture notes. All students in the university used the same LMS and its basic functionalities. Also it is assumed that, all the students had similar pre-knowledge to use this system since, there is a must ICT course for all students in the university taken the first semester and the system was introduced to them in this course. The survey instrument was distributed to the participants in Turkish (the native language of the users) through a link to the survey sent to students’ school e-mail accounts and it was administered online over a period of three months.

Sample

The questionnaire was sent to 470 students. After data collection, null, incomplete and repetitive scores were removed and 253 complete responses were included in analysis. The sample consisted of 57.3% female and 42.7% male students. The age of the participants ranged from 19 to 40 with the mean age being 23.45. The participants ranged from freshmen to PhD students.

The study was conducted with students from six different educational areas. The percentages of the participants by educational area were 40.3% for educational sciences, 23% for engineering sciences, 14% for art and sciences, 10.9% for interdisciplinary sciences and 3.2% for architecture. Furthermore, 53.2% of the students had been using computer for more than 10 years and 75% reported to have good computers skills. Concerning other LMSs, 18% of the participants were familiar with Moodle, 13% with Blackboard and 4% with WebCT. Lastly, the participants were asked whether they were willing to use LMS to support traditional courses and 88% responded positively.

Figure 2. Home page of NET-ClassR
Data analysis and findings

The data was prepared for further analyses by detecting outliers, conducting a missing value analysis, valuating multicollinearity analysis and checking the normality assumption. Firstly, the outliers and their effect on the dataset were analyzed by comparing the mean and trimmed mean values (Walfish, 2006). The difference between these two values was not high; therefore, there was no problematic outlier value in the dataset. Secondly, since the missing values in the dataset did not exceed 10% (Hair, Black, Babin, Anderson, & Tatham, 2006), they were handled using the mean substitution method. Thirdly, VIF values were less than 5 (Hair et al., 2006) indicating that there was no multicollinearity issue between the interaction factors. Lastly, the normality assumption was evaluated with the Kolmogorov–Smirnov test (Field, 2009). According to the results, all the items were found to be significant ($p < .05$). In addition, the skewness and kurtosis ($>-1$ or $<+1$) were analyzed (Huck, 2000) and some problematic items were detected. According to the results, data was not normally distributed.

Factor and reliability analysis

The factor structure of the dataset was examined using an exploratory factor analysis (Stevens, 2012), which was conducted together with the principal axis factors extraction method since the assumption of multivariate normality is violated (Fabrigar, Wegener, MacCallum & Strahan, 1999). As rotation method direct oblimin was selected since the scale items were correlated (Field, 2009). Kaiser-Meyer-Olkin was found to be 0.941, which is higher than the minimum sample size required for factor analysis (0.5) (Field, 2009). In addition, Bartlett’s test of sphericity values were $X^2(946) = 8001.115$ ($p < .001$), which indicated that the dataset provided a meaningful factor structure.

With the exploratory factor analysis, seven different factors were obtained explaining 66.74% of total variance. In contrary to hypotheses, which proposed eight constructs, exploratory factor analysis released seven constructs. Each measurement item of SE was clustered under the PEOU factor. Table 3 presents the new factor structure, related factor loadings (FL) and Cronbach’s alpha of each factor. Seven constructs were found to be reliable having alpha values greater than the required score of 0.7 (Hair et al., 2006). In addition, the overall questionnaire was significantly reliable with an alpha value of 0.96.
Table 3. Factor analysis and reliability results

<table>
<thead>
<tr>
<th>Item ID</th>
<th>New item ID</th>
<th>Factor loading</th>
<th>Factor loading</th>
<th>Factor loading</th>
<th>Factor loading</th>
<th>Cronbach Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item1</td>
<td>PU1</td>
<td>-.846</td>
<td>.887</td>
<td>.887</td>
<td>.887</td>
<td>.887</td>
</tr>
<tr>
<td>Item6</td>
<td>PU2</td>
<td>-.701</td>
<td>.701</td>
<td>.701</td>
<td>.701</td>
<td>.701</td>
</tr>
<tr>
<td>Item44</td>
<td>PU3</td>
<td>-.501</td>
<td>.501</td>
<td>.501</td>
<td>.501</td>
<td>.501</td>
</tr>
<tr>
<td>Item39</td>
<td>PU4</td>
<td>-.347</td>
<td>.347</td>
<td>.347</td>
<td>.347</td>
<td>.347</td>
</tr>
<tr>
<td>Item3</td>
<td>PU5</td>
<td>-.334</td>
<td>.334</td>
<td>.334</td>
<td>.334</td>
<td>.334</td>
</tr>
<tr>
<td>Item9</td>
<td>PEOU1</td>
<td>-.828</td>
<td>.828</td>
<td>.828</td>
<td>.828</td>
<td>.828</td>
</tr>
<tr>
<td>Item18</td>
<td>PEOU2</td>
<td>-.802</td>
<td>.802</td>
<td>.802</td>
<td>.802</td>
<td>.802</td>
</tr>
<tr>
<td>Item20</td>
<td>PEOU3</td>
<td>-.802</td>
<td>.802</td>
<td>.802</td>
<td>.802</td>
<td>.802</td>
</tr>
<tr>
<td>Item8</td>
<td>PEOU4</td>
<td>-.750</td>
<td>.750</td>
<td>.750</td>
<td>.750</td>
<td>.750</td>
</tr>
<tr>
<td>Item17</td>
<td>PEOU5</td>
<td>-.704</td>
<td>.704</td>
<td>.704</td>
<td>.704</td>
<td>.704</td>
</tr>
<tr>
<td>Item10</td>
<td>PEOU6</td>
<td>-.631</td>
<td>.631</td>
<td>.631</td>
<td>.631</td>
<td>.631</td>
</tr>
<tr>
<td>Item7</td>
<td>PEOU7</td>
<td>-.616</td>
<td>.616</td>
<td>.616</td>
<td>.616</td>
<td>.616</td>
</tr>
<tr>
<td>Item21</td>
<td>PEOU8</td>
<td>-.516</td>
<td>.516</td>
<td>.516</td>
<td>.516</td>
<td>.516</td>
</tr>
<tr>
<td>Item19</td>
<td>PEOU9</td>
<td>-.503</td>
<td>.503</td>
<td>.503</td>
<td>.503</td>
<td>.503</td>
</tr>
<tr>
<td>Item11</td>
<td>PEOU10</td>
<td>-.349</td>
<td>.349</td>
<td>.349</td>
<td>.349</td>
<td>.349</td>
</tr>
<tr>
<td>Item12</td>
<td>BI1</td>
<td>.856</td>
<td>.856</td>
<td>.856</td>
<td>.856</td>
<td>.856</td>
</tr>
<tr>
<td>Item15</td>
<td>BI2</td>
<td>.787</td>
<td>.787</td>
<td>.787</td>
<td>.787</td>
<td>.787</td>
</tr>
<tr>
<td>Item16</td>
<td>BI3</td>
<td>.807</td>
<td>.807</td>
<td>.807</td>
<td>.807</td>
<td>.807</td>
</tr>
<tr>
<td>Item14</td>
<td>BI4</td>
<td>.736</td>
<td>.736</td>
<td>.736</td>
<td>.736</td>
<td>.736</td>
</tr>
<tr>
<td>Item2</td>
<td>BI5</td>
<td>.649</td>
<td>.649</td>
<td>.649</td>
<td>.649</td>
<td>.649</td>
</tr>
<tr>
<td>Item13</td>
<td>BI6</td>
<td>.652</td>
<td>.652</td>
<td>.652</td>
<td>.652</td>
<td>.652</td>
</tr>
<tr>
<td>Item4</td>
<td>BI7</td>
<td>.547</td>
<td>.547</td>
<td>.547</td>
<td>.547</td>
<td>.547</td>
</tr>
<tr>
<td>Item5</td>
<td>BI8</td>
<td>.496</td>
<td>.496</td>
<td>.496</td>
<td>.496</td>
<td>.496</td>
</tr>
<tr>
<td>Item36</td>
<td>BI9</td>
<td>.481</td>
<td>.481</td>
<td>.481</td>
<td>.481</td>
<td>.481</td>
</tr>
<tr>
<td>Item22</td>
<td>ENJ1</td>
<td>.830</td>
<td>.830</td>
<td>.830</td>
<td>.830</td>
<td>.830</td>
</tr>
<tr>
<td>Item25</td>
<td>ENJ2</td>
<td>.790</td>
<td>.790</td>
<td>.790</td>
<td>.790</td>
<td>.790</td>
</tr>
<tr>
<td>Item26</td>
<td>ENJ3</td>
<td>.802</td>
<td>.802</td>
<td>.802</td>
<td>.802</td>
<td>.802</td>
</tr>
<tr>
<td>Item23</td>
<td>ENJ4</td>
<td>.714</td>
<td>.714</td>
<td>.714</td>
<td>.714</td>
<td>.714</td>
</tr>
<tr>
<td>Item24</td>
<td>ENJ5</td>
<td>.628</td>
<td>.628</td>
<td>.628</td>
<td>.628</td>
<td>.628</td>
</tr>
<tr>
<td>Item30</td>
<td>ENJ6</td>
<td>.390</td>
<td>.390</td>
<td>.390</td>
<td>.390</td>
<td>.390</td>
</tr>
<tr>
<td>Item28</td>
<td>SN1</td>
<td>.780</td>
<td>.780</td>
<td>.780</td>
<td>.780</td>
<td>.780</td>
</tr>
<tr>
<td>Item33</td>
<td>SN2</td>
<td>.764</td>
<td>.764</td>
<td>.764</td>
<td>.764</td>
<td>.764</td>
</tr>
<tr>
<td>Item31</td>
<td>SN3</td>
<td>.630</td>
<td>.630</td>
<td>.630</td>
<td>.630</td>
<td>.630</td>
</tr>
<tr>
<td>Item29</td>
<td>SN4</td>
<td>.418</td>
<td>.418</td>
<td>.418</td>
<td>.418</td>
<td>.418</td>
</tr>
<tr>
<td>Item27</td>
<td>SN5</td>
<td>.338</td>
<td>.338</td>
<td>.338</td>
<td>.338</td>
<td>.338</td>
</tr>
<tr>
<td>Item32</td>
<td>SN6</td>
<td>.384</td>
<td>.384</td>
<td>.384</td>
<td>.384</td>
<td>.384</td>
</tr>
<tr>
<td>Item34</td>
<td>STS1</td>
<td>-.433</td>
<td>.433</td>
<td>.433</td>
<td>.433</td>
<td>.433</td>
</tr>
<tr>
<td>Item37</td>
<td>STS2</td>
<td>-.343</td>
<td>.343</td>
<td>.343</td>
<td>.343</td>
<td>.343</td>
</tr>
<tr>
<td>Item38</td>
<td>STS3</td>
<td>-.344</td>
<td>.344</td>
<td>.344</td>
<td>.344</td>
<td>.344</td>
</tr>
<tr>
<td>Item35</td>
<td>STS4</td>
<td>-.303</td>
<td>.303</td>
<td>.303</td>
<td>.303</td>
<td>.303</td>
</tr>
<tr>
<td>Item41</td>
<td>IC1</td>
<td>.686</td>
<td>.686</td>
<td>.686</td>
<td>.686</td>
<td>.686</td>
</tr>
<tr>
<td>Item42</td>
<td>IC2</td>
<td>.704</td>
<td>.704</td>
<td>.704</td>
<td>.704</td>
<td>.704</td>
</tr>
<tr>
<td>Item40</td>
<td>IC3</td>
<td>.631</td>
<td>.631</td>
<td>.631</td>
<td>.631</td>
<td>.631</td>
</tr>
<tr>
<td>Item43</td>
<td>IC4</td>
<td>.437</td>
<td>.437</td>
<td>.437</td>
<td>.437</td>
<td>.437</td>
</tr>
</tbody>
</table>

Extraction Method: Principal Axis Factoring.
Rotation Method: Oblimin with Kaiser Normalization. a
a. Rotation converged in 15 iterations.

Model assessment

The dataset did not follow a multivariate normal distribution; therefore, the proposed model was assessed with component-based structural equation modelling, specifically partial least square (PLS) (Chin, 1998) using SmartPLS software. PLS was used since it is a method suitable to cases when relationships among theoretical constructs are explored and overall nomological network has not been well understood (Peng & Lai, 2012). Before the evaluation of the structural model, sample size requirement and preliminary data analysis including outlier detection, missing value analysis, multicollinearity analysis and normality checks were performed (Hair.
et al., 2006). “10 times” rule of thumb (Peng & Lai, 2012) was used for sample size requirement, in which our sample size of 253 was adequate to conduct the analysis. The proposed research model was verified through measurement and structural assessment.

**Measurement model**

Confirmatory factor analysis (CFA) was employed to assess the measurement model in terms of convergent validity and discriminant validity. Convergent validity was assessed by FL, Average Variance Extracted (AVE) and Composite Reliability (CR) (Table 4). Each observed variable must load its latent variable with at least 0.7 to provide adequate convergent validity (Hair et al., 2006). PEOU9, PEOU10 and ENJ6 did not have an adequate load on the related latent variables and therefore they were extracted from the dataset. Since the loadings of PEOU4, SN1 and SN5 were only slightly lower than 0.7, they were not excluded. For internal consistency, the AVE value should be higher than 0.5 and CR value should be 0.7 or higher for each latent variable (Hair et al., 2006). Considering the AVE and CR values, the dataset had adequate convergent validity.

**Table 4. Convergent validity results**

<table>
<thead>
<tr>
<th>Item ID</th>
<th>Factor Loadings</th>
<th>Composite Reliability</th>
<th>AVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>PU1</td>
<td>.828</td>
<td>.916</td>
<td>68%</td>
</tr>
<tr>
<td>PU2</td>
<td>.832</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PU3</td>
<td>.802</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PU4</td>
<td>.822</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PU5</td>
<td>.864</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEOU1</td>
<td>.825</td>
<td>.926</td>
<td>61%</td>
</tr>
<tr>
<td>PEOU2</td>
<td>.822</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEOU3</td>
<td>.854</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEOU4</td>
<td>.658</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEOU5</td>
<td>.723</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEOU6</td>
<td>.722</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEOU7</td>
<td>.776</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEOU8</td>
<td>.801</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BI1</td>
<td>.896</td>
<td>.956</td>
<td>70%</td>
</tr>
<tr>
<td>BI2</td>
<td>.862</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BI3</td>
<td>.872</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BI4</td>
<td>.825</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BI5</td>
<td>.874</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BI6</td>
<td>.850</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BI7</td>
<td>.797</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BI8</td>
<td>.789</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BI9</td>
<td>.801</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENJ1</td>
<td>.903</td>
<td>.931</td>
<td>73%</td>
</tr>
<tr>
<td>ENJ2</td>
<td>.876</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENJ3</td>
<td>.749</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENJ4</td>
<td>.889</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENJ5</td>
<td>.853</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SN1</td>
<td>.602</td>
<td>.856</td>
<td>54%</td>
</tr>
<tr>
<td>SN2</td>
<td>.710</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SN3</td>
<td>.735</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SN4</td>
<td>.751</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SN5</td>
<td>.687</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SN6</td>
<td>.747</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STS1</td>
<td>.825</td>
<td>.897</td>
<td>68%</td>
</tr>
<tr>
<td>STS2</td>
<td>.864</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STS3</td>
<td>.880</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STS4</td>
<td>.794</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IC1</td>
<td>.876</td>
<td>.886</td>
<td>66%</td>
</tr>
<tr>
<td>IC2</td>
<td>.810</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IC3</td>
<td>.764</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IC4</td>
<td>.801</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The discriminant validity refers that a measure does not correlate highly with another measure (Peter, 1981). In order to prove discriminant validity, the square root of the AVE values for each construct on the diagonal should be higher than the correlations with the related construct and all other correlations (Peter, 1981). Table 5 shows that square root of AVE for each construct on the diagonal was greater than the other values. Therefore, it can be concluded that the constructs of the dataset were adequately different from each other.

Table 5. Discriminant validity

<table>
<thead>
<tr>
<th></th>
<th>BI</th>
<th>ENJ</th>
<th>IC</th>
<th>PEOU</th>
<th>PU</th>
<th>SN</th>
<th>STS</th>
</tr>
</thead>
<tbody>
<tr>
<td>BI</td>
<td>0.8417</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENJ</td>
<td>0.58064</td>
<td>0.85565</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IC</td>
<td>0.573108</td>
<td>0.537491</td>
<td>0.813754</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEOU</td>
<td>0.669152</td>
<td>0.38486</td>
<td>0.391677</td>
<td>0.78381631</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PU</td>
<td>0.729679</td>
<td>0.654968</td>
<td>0.713007</td>
<td>0.495454</td>
<td>0.82977</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SN</td>
<td>0.463768</td>
<td>0.411048</td>
<td>0.514843</td>
<td>0.264658</td>
<td>0.535481</td>
<td>0.73714779</td>
<td></td>
</tr>
<tr>
<td>STS</td>
<td>0.662742</td>
<td>0.576064</td>
<td>0.641242</td>
<td>0.512061</td>
<td>0.699444</td>
<td>0.54901</td>
<td>0.8287</td>
</tr>
</tbody>
</table>

Structural model

The structure of the proposed research model was examined by considering the path coefficient values to assess the statistical significance of each hypothesis. The dataset containing 253 samples was analyzed following a bootstrapping procedure and the significance of difference between the constructs was evaluated. Figure 4 presents the estimated path coefficients.

According to the results of the structural model (Table 6), none of the measurement items were clustered under SE; therefore, the model was assessed by extracting this construct, and H11 and H12 could not be evaluated.

Except for H10, which examined the relationship between SN and BI, all the other hypotheses were accepted. A strong positive relationship was found between the constructs of H1, H2, H3, H5, H6, H9 and H13 at the level of \( p < .001 \). In addition, PEOU and BI were related at \( p < .001 \), which had not been initially hypothesized. The relationships between the constructs H7 and H8 were also significant at \( p < .01 \). Finally, H4 was supported and found to be significant at \( p < .05 \).
### Table 6. Summary of hypotheses tests

<table>
<thead>
<tr>
<th>H</th>
<th>Relationships</th>
<th>t-Values</th>
<th>β</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>PU -&gt; STS</td>
<td>9.270</td>
<td>0.564***</td>
<td>Accepted</td>
</tr>
<tr>
<td>H2</td>
<td>PU -&gt; BI</td>
<td>5.872</td>
<td>0.351***</td>
<td>Accepted</td>
</tr>
<tr>
<td>H3</td>
<td>PEOU -&gt; PU</td>
<td>4.101</td>
<td>0.180***</td>
<td>Accepted</td>
</tr>
<tr>
<td>H4</td>
<td>STS -&gt; BI</td>
<td>2.186</td>
<td>0.133*</td>
<td>Accepted</td>
</tr>
<tr>
<td>H5</td>
<td>ENJ -&gt; PU</td>
<td>5.384</td>
<td>0.308***</td>
<td>Accepted</td>
</tr>
<tr>
<td>H6</td>
<td>ENJ -&gt; PEOU</td>
<td>4.058</td>
<td>0.245***</td>
<td>Accepted</td>
</tr>
<tr>
<td>H7</td>
<td>ENJ -&gt; STS</td>
<td>3.219</td>
<td>0.207**</td>
<td>Accepted</td>
</tr>
<tr>
<td>H8</td>
<td>ENJ -&gt; BI</td>
<td>2.627</td>
<td>0.107**</td>
<td>Accepted</td>
</tr>
<tr>
<td>H9</td>
<td>SN -&gt; PU</td>
<td>3.350</td>
<td>0.157***</td>
<td>Accepted</td>
</tr>
<tr>
<td>H10</td>
<td>SN -&gt; BI</td>
<td>1.570</td>
<td>0.061</td>
<td>Rejected</td>
</tr>
<tr>
<td>H11</td>
<td>SE -&gt; PEOU</td>
<td>-</td>
<td>-</td>
<td>Rejected</td>
</tr>
<tr>
<td>H12</td>
<td>SE -&gt; BI</td>
<td>-</td>
<td>-</td>
<td>Rejected</td>
</tr>
<tr>
<td>H13</td>
<td>IC -&gt; PU</td>
<td>7.000</td>
<td>0.396***</td>
<td>Accepted</td>
</tr>
<tr>
<td>HAD</td>
<td>PEOU -&gt; BI</td>
<td>5.825</td>
<td>0.370***</td>
<td>Accepted</td>
</tr>
</tbody>
</table>

Note. *p < .05; **p < .01; ***p < .001; HAD: Additional Hypothesis.

### Discussion

This research was conducted to examine the factors that affected students’ behavioral intention towards LMS use in higher education. The constructs of TAM (perceived usefulness, perceived ease-of-use and behavioral intention) were taken as the starting point of the proposed research model. TAM was extended by adding external factors to predict the constructs of original TAM; namely, satisfaction, enjoyment, subjective norm, and interactivity and control. The relationships between these constructs were analyzed using structural equation modeling.

#### Perceived usefulness and perceived ease-of-use

The results revealed that perceived usefulness and perceived ease-of-use are significant predictors of behavioral intention towards LMS use. When the relationship between these two predictors was examined, students’ perception of usefulness was found to directly and significantly affect students’ behavioral intention towards LMS use. This relationship implies that when the students perceive the system to be useful, their behavioral intention to use the system increases. This finding validates the findings of the previous studies conducted by Lee et al. (2005), Saadé and Bahli (2005), and Yi and Hwang (2003). E-learning systems should be designed and developed to add value to student learning and the value of these systems can be improved by providing enhanced e-learning services (Lee et al., 2009).

A positive relationship was found between perceived ease-of-use and perceived usefulness. This finding showed that perceived ease-of-use significantly affects perceived usefulness, which means that if students consider it easy to use an LMS, they feel that using an e-learning system is more useful. Similarly, Lee et al. (2009) reported that perceived ease-of-use is a significant antecedent of perceived usefulness and the design of learning content is important for increasing easiness perception. In addition, this finding has a significance in terms of designing systems with low complexity to improve the value of e-learning services (Lee et al., 2009). In addition to this relationship, a positive relationship was found between perceived ease-of-use and behavioral intention, which had not been hypothesized in the proposed research model. This relationship implies that when the users of the system perceive that the system is easy to use, their behavioral intention to use the system increases. This finding is supported by Lee (2008), who suggested that a system should be developed to target changes in perceived ease-of-use to increase students’ adoption of online learning systems.

#### Enjoyment

The results showed that enjoyment is another significant predictor of student’s intention towards LMS use. It also has significant relationships with the constructs of perceived ease-of-use, perceived usefulness and satisfaction. These results are similar to those reported by Yi and Hwang (2003) indicating that students’ perceived enjoyment has an important effect on their perception of the usefulness and easiness of LMS. Moreover, users’ perceived enjoyment has a more effective role than users’ perceived ease-of-use in determining...
students’ perceived usefulness of the system. In addition to these relationships, it was observed that enjoyment had a significant and positive effect on satisfaction. This implies that if students enjoy and have fun throughout the interaction with LMS and using the e-learning services, they will be more willing to use it in the future. This finding is in parallel with the results of previous studies. Sørebø, Halvari, Gulli, and Kristiansen (2009) examined enjoyment as an intrinsic motivation and found a significant relation between intrinsic motivation and satisfaction for using e-learning technology. In addition, in the current study, enjoyment was found to have a significant effect on behavioral intention, which supports the findings of Lee et al. (2005). The researchers found that perceived enjoyment might be the key element for the adoption and use of internet-based learning media. Therefore, instructors should create a learning environment by considering content variation, fun creation, immediate feedback and interaction encouragement issues to increase the use of online learning environments.

**Subjective norm**

The results of the study showed that subjective norm significantly affects the students’ perceived usefulness of LMS in the higher education context. Similarly, Park (2009) found a significant relationship between subjective norm and perceived usefulness. The researcher provided one possible explanation for this relationship: Subjective norm is an extrinsic motivational factor that could help university students self-regulate their motivation on e-learning (Park, 2009). In the current study, in contrast to this finding of Park (2009), subjective norm was not found to be a predictor of behavioral intention towards LMS use. This may have resulted from the participant students being obliged to use the system. Therefore, their intention may not have been affected by their social environment.

**Satisfaction**

The results revealed that satisfaction is a significant predictor of students’ behavioral intention towards LMS use in higher education. A positive and significant relationship was found between satisfaction and behavioral intention towards LMS use. This relationship implies that when the users are satisfied with using LMS, their behavioral intention toward LMS use is affected positively for future use. Similarly, Roca, Chiu, and Martínez (2006), and Lee (2010) found that satisfaction positively affects continuance intention to use e-learning applications.

**Interactivity and control**

In the current study, the relationship between perceived usefulness and interactivity and control was also examined. The results showed that the interactivity and control construct has a significant direct effect on the participants’ perceived usefulness, which implies that this construct can determine students’ perceived usefulness of LMS. In addition, students’ perception of the usefulness of LMS increased after they used this system to have more control over their learning process. Therefore, e-learning systems and services should support interactivity and control by supporting communication between the instructor and students, offering tools such as chat, forum and e-mail to strengthen their relationship and providing an environment for students to learn at their own pace.

In this study, the effect of self-efficacy on users’ perceived ease-of-use and behavioral intention could not be examined since none of the measurement items were clustered under the self-efficacy factor. However, the research available in the literature has already demonstrated that application-specific self-efficacy has a significant effect on the behavioral intention of the system’s users (Yi & Hwang, 2003) and is more powerful than behavioral intention in determining the actual use of the system. In addition, Park (2009) reported that self-efficacy plays an important role in affecting attitude towards e-learning and behavioral intention to use e-learning. However, further studies are needed to examine self-efficacy with a new sample and new measurement items to reveal its effect on students’ future intention of LMS use.

In brief, it was observed that the validity measures of the research model were effective in predicting the behavioral intention of the participants. The research model explained 68% of the behavioral intention of students towards LMS use. This result also provides a reliable prediction about students’ behavioral intention towards LMS use in future.
Limitations and implications for future research

We consider this research to be a valuable guideline for researchers who will undertake research on the acceptance of LMS in the e-learning context. However, the current study has certain limitations. First, the data was collected from the students of the same university, which affected the representativeness of the sample and the generalization of the results. The range of universities and the sample size of the students using this system should be extended to improve the generalizability of the results. In addition, the measurement items of self-efficacy did not load under one factor; therefore, this construct and its proposed relations could not be analyzed. This construct should be re-assessed with a new sample. Moreover, quantitative research methodology was applied in this study. However, an in-depth qualitative examination would reveal personal opinions and detailed reasons explaining the relationships between the proposed constructs. Therefore, further studies should support their quantitative findings with a qualitative approach. Finally, the model should be extended with an additional variable to improve the model’s prediction power to account for the remaining 32% of user intentions. In addition, further studies with cross-sectional and cross-cultural approaches are required to increase the predictive value of LMS-TAM.

Conclusion

This study examined the factors affecting students’ behavioral intention towards LMS use in the higher education context based on quantitative research. A structural research model was proposed and validated through an online survey. LMS-TAM is an extended version of TAM including the external factors of enjoyment, subjective norm, satisfaction and interactivity and control. In addition to perceived ease of use and perceived usefulness effects over behavioral intention validated in original TAM, LMS-TAM implies that users’ behavioral intention is influenced directly or indirectly by enjoyment, subjective norm, satisfaction and interactivity and control factors. Systems increasing students’ enjoyment, satisfaction and interactivity and control are more tended to be accepted by the students. The factors measuring students’ behavioral intention towards LMS use included in LMS-TAM model are not directly related to specific functions of the LMS, they are related to students’ general perceptions. The LMS used in this study, Moodle, and the other LMSs provides similar features to the end users, despite they have different user interfaces. Therefore, we believe that the results will be helpful to improve different type of LMSs and increase their usage. In this context, this study contributes to the related literature by developing a new model for students’ intention towards LMS use. LMS-TAM has potential to be a predictive model for studies on students’ acceptance of e-learning.

Acknowledgments

This research was supported by the Scientific and Technological Research Council of Turkey, under project number 109K394.

References


Instructional Suggestions Supporting Science Learning in Digital Environments Based on a Review of Eye Tracking Studies

Fang-Ying Yang1*, Meng-Jung Tsai2, Guo-Li Chiou1, Silvia Wen-Yu Lee3, Cheng-Chieh Chang4 and Li-Ling Chen1

1Graduate Institute of Science Education, National Taiwan Normal University, Taipei City, Taiwan // 2Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taipei City, Taiwan // 3Graduate Institute of Science Education, National Changhua University of Education, Changhua City, Taiwan // 4Center of Teacher Education, National Taiwan Ocean University, Keelung, Taiwan

*Corresponding author

(Submitted October 23, 2016; Revised December 4, 2016; Accepted January 11, 2017)

ABSTRACT

The main purpose of this study was to provide instructional suggestions for supporting science learning in digital environments based on a review of eye tracking studies in e-learning related areas. Thirty-three eye-tracking studies from 2005 to 2014 were selected from the Social Science Citation Index (SSCI) database for review. Through a literature analysis program, CATA, and in-depth content analysis of the research methods and findings, five research theme clusters were abstracted from the selected papers, namely, cognitive activities in processing multimedia presentations, multimedia effects, roles of personal factors, effects of instructional design, and learning with dynamic e-platforms. Based on the results of the in-depth theme analyses, it is suggested that the design of e-learning instruction should consider placing related text and graphics in adjacent areas, using one verbal mode at a time, and providing explicit and clear verbal explanations. When using animations, instructors need to explain carefully the goals and contents of the animations to reduce the extraneous cognitive load. In the dynamic learning environment, a pre-training program is necessary for students to become familiar with the new environment. Finally, individual differences such as background knowledge, cognitive abilities and cognitive styles should be taken into consideration in the instructional design.

Keywords

Instructional design, Eye tracking, Digital learning, Digital learning environment, Science learning

Introduction

With the rapid development of digital technologies such as the Internet, multimedia and mobile devices, the conventional learning environments with the blackboard as the main presentation platform have gradually evolved into electronic classrooms. The infusion of technologies into teaching and learning has given rise to new paradigms of learning, including multimedia learning, web-based inquiry learning, computer supported collaborative learning, game-based learning, e-learning, mobile learning and so forth (e.g., Mayer, 2014; Slotta, Linn, & Lee, 2009; Lai, Hwang, Liang, & Tsai, 2016). It is believed that technologies function as tools to support knowledge construction, as information vehicles for exploring knowledge, as media to support learning-by-doing, as social media to support cooperative learning and communication, and as intellectual partners to support learning-by-reflecting (Jonassen, Peck, & Wilson, 1999; Goodyear & Retalis, 2010; Hwang & Wang, 2016). Practically, the use of technologies is more prevalent in science classrooms due to the close link between the process of knowledge construction in science and the use of technologies. In the science education literature, significant numbers of studies regarding digital learning have been accumulated (e.g., Lee, Tsai, & Wu, 2011; Tang & Tsai, 2016).

Although numerous studies related to the use of technologies in classrooms have been reported, there are scholars who have questioned the effectiveness of the use of technologies in assisting teaching and learning (e.g., West, Wright, Gabbitas, & Graham, 2006; Cheung & Slavin, 2013; Noesgaard & Orngreen, 2015). Indeed, because of their rapid development, new technologies are often applied to classrooms without in-depth examination of their impacts on learning before implementation. To address this issue, researchers who study the use of digital technologies in classrooms are paying attention to not only whether these technologies can improve students’ learning performance, but also to how they can assist student learning (e.g., Chang & Yang, 2010; Groff, Howells, & Cranmer, 2012). To explore the latter issues, methods that can reveal how students learn with technologies are preferred. Surveys and content analysis of video data, computer log files, or interview data are frequently employed for that purpose (e.g., Groff, Howells, & Cranmer, 2012; Ryoo & Linn, 2014; Shirley &
Irving, 2015). Although the above-mentioned data types are records of learning behaviors, they are limited in terms of revealing the real-time actions and changes of mind during learning.

A promising method capable of recording online cognitive activities is the eye tracking method. Based on the eye-mind hypothesis (Just & Carpenter, 1980) that what is fixated on is processed, the eye tracking method has been developed to study mechanisms of information processing and a variety of cognitive processes (e.g., Rayner, 2009; Gidlöf, Wallin, Dewhurst, & Holmqvist, 2013). In addition to cognitive activities, the method is also applied to research areas such as human-computer interaction and usability research (Jocab & Karn, 2003; McEwen & Dubé, 2015). In eye tracking studies, researchers use eye tracking measures to indicate learners’ or users’ visual attention, workload or mental effort expended during a task. Fixations and saccades are the two major eye movements identified by eye trackers. The former refers to a relatively stable state of eye movement, while the latter refers to the rapid movement between any two consecutive fixations (Rayner, 2009). Eye tracking measures, which are generally categorized according to temporal, spatial and frequency scales (Lai et al., 2013), are usually analyzed and reported based on areas of interest (AOIs) as defined in individual studies. Depending on the research questions, eye tracking measures are selected or analyzed differently with respect to different study purposes (Holmqvist et al., 2011).

The use of eye tracking methods in educational studies has proliferated only in recent years. According to Lai et al. (2013), before 2009, few educational studies had used such methods, probably due to the lack of communication between the fields of education and psychology, and the high cost of the equipment. The review study by Lai and colleagues found that two major issues discussed by the educational researchers who applied eye tracking methods in their studies are patterns of information processing and effects of instructional designs. It is interesting to note that in the work of Lai et al. (2013), quite a few papers concerning the instructional designs are related to multimedia learning. Nevertheless, how the eye tracking method can reveal the use of technologies in assisting science learning has not been systematically analyzed. Given that the employment of multimedia or digital technologies in the science classroom is an unavoidable trend, we have thus made an attempt to review eye-tracking studies regarding science learning in multimedia as well as technology-enhanced environments (we call such environments, “digital learning environments”). Drawn from the review, we hope to propose suggestions for future instructional design supporting learning in digital learning environments.

**Method**

The empirical papers included in the study came from the Social Science Citation Index (SSCI) database. To select suitable papers, the keywords of “eye tracking” and “eye movement” were combined with “multimedia learning,” “e-learning,” “digital learning,” “learning technology,” “educational technology,” “mobile learning,” and “ubiquitous learning,” using the Boolean operator “AND.” The years spanned from 2005 to 2014. There were 91 papers extracted from the first round of the search process. The same keyword combinations were then put together using the Boolean operator “OR.” Consequently, 46 of the original 91 papers remained. Afterwards, the researchers manually inspected the studies. Those involving non-science subject areas, non-human subjects, subjects with medical impairment or special needs, or which were conducted in out-of-school contexts were excluded from the analysis. In addition, only empirical studies were retained. As a result, a total of 33 papers, including 36 studies, were selected from the SSCI database for this review.

To categorize the characteristics of the selected papers, we used a paper analysis program, Content Analysis Toolkit for Academic Research (CATA), to do the initial grouping. The program, using data mining technology, was developed by Yseng and Tsay (2013), and is able to analyze the similarity of papers in accordance with the keywords, citations, authors and so forth. The quality of CATA has been compared to other similar data mining systems (such as infoMap developed by Computational Semantics Library at Stanford University), and the result suggested that the method is highly reliable (for more details, see Tseng, 2010). Since CATA classifies academic papers in accordance with similarities among papers, problems of subjectivity and uncertainty caused by content analysis alone can be reduced. In other words, instead of researchers alone determining the structure of these selected papers, CATA allows the papers themselves to reveal their structure. The preliminary analysis by CATA suggested that the 33 papers analyzed in the study fell into 7 clusters with 4 unidentified.

After the CATA analysis, the researchers reviewed and examined the paper contents to determine the relevance of the papers in each cluster, and reassigned papers to different clusters if necessary. Among the seven clusters abstracted by CATA, two consisted of only two papers. After examining their contents, they were re-assigned to different clusters. In short, as a triangulation process, CATA and content analysis complemented each other to
help the researchers identify possible research themes embedded in the selected papers. As a result, five research themes as will be introduced later were determined.

After all the selected papers were assigned to their proper categories, they were analyzed in-depth according to their research purposes, subjects, models of eye trackers, learning topics, eye movement measures and indications, instructional materials, experimental design, performance measurements and research results. It should be noted that the learning topics and types of eye movement measures were analyzed based on the categories of learning topics and eye movement measures mentioned in the study of Lai et al. (2013) The review task went through intensive discussions by all the authors of the paper; therefore, there was no issue of inter-coder reliability.

**Result and discussion**

**General characteristics of the reviewed papers**

Among the 36 studies (from 33 papers), the majority \( n = 26 \) recruited research subjects at universities or colleges. In only seven papers were the research subjects K-12 students. The remaining three papers studied adult learners, including university students, medical professionals, and surgical residents. In terms of learning topics, conceptual development received the most attention (27 studies). There are 5 studies focusing on skill or strategy learning, and the rest 2 studies on perception.

In contrast to the reading studies that frequently employ eye-tracking systems with high sampling rates (> 500Hz), eye trackers with sampling rates of 50Hz ad 60Hz were the most popular types of systems mentioned in the reviewed papers. In the searched studies, eye movements were measured using the following indicators: fixation frequency and duration, gaze frequency and duration, saccade frequency and duration, scan path, reading time, and transitions between different representations. We further categorized the eye movement indicators into temporal, spatial, and count measures. Almost all the papers collected temporal measures of eye movement (34 studies), followed by spatial measures (18 studies) and frequency measures (15 studies). Evidently, temporal eye movement measures were used in almost all studies to indicate the time needed to process information, and those that revealed the attention allocations and spatial sequences, such as the number of fixations on different AOIs, regression counts, and scan paths, were frequently reported.

The above-mentioned eye movement measures reflect learners' online cognitive activities that should explain consequent learning behaviors or performances. In most of the reviewed papers, the behavior or performance measurements included, in general, three major types, namely, achievement test, skill/ability test, and performance exam. Achievement tests included prior knowledge tests, comprehension tests, delayed/retention tests, and transfer tests. Several studies utilized tests to examine students' skills or abilities such as visualization tests, spatial ability tests, and visual memory tests. Performance exams, such as knot tying performance and diagnostic performance, were conducted in the studies on medical knowledge. Other qualitative data were collected through think-aloud when completing the tasks, verbal responses to questions, and interviews.

**Research themes abstracted from the study findings**

As mentioned, the researchers of this study cross-examined the result of CATA analysis. Subsequently, five clusters of research theme were abstracted, illustrated as follows.

**Cluster I – Cognitive activities for processing multimedia presentations**

Cluster one discusses the information processing activities for different forms and combinations of representations. There are ten papers in this cluster. The learning topics of these papers focus on conceptual development, with one exception that dealt with strategy learning (#19). Participating subjects included university students (6 studies), high school students (two studies) and elementary learners (one study). Four studies used text-based material with text-picture presentations as the learning material (#9, #13, #16, # 28, & #33), three involved visual representations such as image, animation or simulation (#2, #19 & #20), and the remaining three (#18, #21, & #33) examined the reading of WebPages. The multimedia representations included text, graphics, animation, simulation, and narration. Eye trackers with sampling rates of 60Hz were the most popular system. As for the eye movement measures, fixation related measures were reported most frequently,
showing the processing time and attention locations as well as distributions. One study (#19) found that the fixation duration corresponded to the use of problem-solving strategies. Scan paths (i.e., saccades or transitions between different interest areas) were used to indicate the information integration strategies. Three studies (#16, #18, & #33) employed the measure of pupil dilation, while another used the fixation duration and saccade length to indicate cognitive load.

Several important findings can be summarized from the studies in this cluster. It was found that when written text and pictures were presented together, readers attended more to the written text than to the graphics. Nevertheless, when the graphics were in a conceptual, sequential or interactive form, the visual attention and processing time of the target areas increased. It was found that spoken text and background knowledge played a role in guiding visual attention. While spoken text seemed to increase visual attention to and processing time of the interest areas, background knowledge was found to reduce the processing time. When spoken text and written texts with the same content appeared together, the cognitive load increased. Meanwhile, animations, while attracting readers’ attention, might result in extraneous cognitive load. For concept learning, these studies consented that the reading of texts or graphics alone could not predict the concept achievements. Instead, it was the information integration strategies, such as scanning or transitions between different representations, which were related most to the concept achievements. Additionally, an initial glance at the picture seemed to provide a mental scaffold for text reading. In short, the studies in this cluster suggest that learning from multimedia presentations requires learners to strategically distribute their attention to the interest areas and integrate information from different types of representation.

Cluster 2 – Multimedia effects

The theme of Cluster 2 is centered on examining multimedia effects as suggested by the cognitive load theory (Chandler & Sweller, 1992) and the theory of multimedia learning (Moreno & Mayer, 1999). There are five papers (#7, #8, #15, #24, and #26) in this cluster, in which eight studies were involved. The various multimedia effects include the visualization (multimedia representation) effect, the seductive details effect (coherence effect) (#7), the spatial contiguity effect (#15, #26), and the text-modality (split-attention) effect (#8, #24, #26). Specifically, the seductive details effect discusses how interesting but unimportant text and illustrations affect learning performance. The visualization effect discusses how decorative and instructional pictures paired with text affect learners’ distribution of attention. The spatial contiguity effect discusses how integrated presentation (printed words close to corresponding graphics) and separate presentation (printed words at the bottom of the screen) affect learners’ cognitive processing. The text modality effect discusses whether written text and spoken text affect learners’ viewing behaviors and learning outcomes. Among the eight studies, except for one study that involved 7th and 8th graders (#8), all participants were university or graduate students. The eye-trackers employed by these studies had sampling rates ranging from 50Hz to 250Hz. Regarding the eye movement measures involved, all of these eight studies analyzed the spatial measures (e.g., number of switches (regressions) or transitions between text and visualizations) as well as the frequently observed temporal measures (e.g., gaze duration and viewing time), indicating that a great deal of attention was paid to how multimedia presentations may interact with information integration behaviors.

Analysis on the findings of these papers showed that seductive text passages and illustrations were found to hinder the transfer performance differently, but not the retention performance (#7). Pictures that are decorative by nature were found to be neither harmful to nor beneficial for learning. However, it was found that decorative pictures moderated the beneficial effect of instructional pictures for learners with low prior knowledge (#8). Learners’ eye movement patterns suggested that an integrated presentation led to more integrative transitions and corresponding transitions compared to the separate presentation (#15). Visualizations were processed more when coupled with spoken text than with written text (#24, #26). Moreover, when animations were presented, learning performance was related to the reading time spent on the animations (#26).

Cluster 3 – Roles of learner factors

The theme of Cluster 3 constructed from the reviewed studies was centered on comparing the learning behaviors or performances of participants with different backgrounds or preferences. Six studies were included. These studies used the levels of the participants’ prior knowledge (#3, #17, #31), expertise (#12), spatial abilities (#4) or cognitive styles (#30) as the criterion for splitting them into some categorized groups (such as higher/lower prior knowledge), and aimed to compare the differences in learning behaviors or performances of the participants in the different groups. Noticeably, some studies in Clusters 1 and 2 also discussed the background effect as one
of their research questions. However, since the main research issues of these studies were not the effect of background knowledge, they were not assigned to Cluster 3. Four of the six studies used university students as the participants, while the other two used adult learners and high school students, respectively. The sampling rate of all six eye-trackers used in this cluster ranged from 50Hz to 60Hz. With respect to the eye movement measures, five of the six studies used the temporal measures (such as fixation duration and total reading time) to refer to the length of time and amount of attention that the participants used to process the information. Three of the studies used counting measures (such as number of fixations and inter-scanning count) to reveal how the participants allocated their visual attention and how they devoted their mental efforts to integrating information from different sources. Two of the studies used spatial measures (such as scan paths) to identify the patterns of the participants’ visual transitions as a way of exploring their cognitive strategies.

In general, the findings of these six studies confirmed the relationships between some personal factors and eye-movement measures. Participants with higher prior knowledge were more likely to spend more time reading both keywords and graphics, and to make more transitions both between graphics and text and between different graphics (#3, #17, #31). When solving spatial problems, learners with higher spatial test performance displayed better working memory and employed more of the analytic strategies (i.e., more analytic eye movements were found) to solve the problems (#4). In the study on transfer of expertise (#12), it was found that experts allocated their attention to task-relevant areas, indicating success of transfer depending on the similarity of the tasks. As for cognitive style, subjects who were identified as imagers fixated more on images, verbalizers on text, and intermediates equally on images and text. Moreover, a match between cognitive style and learning environment improved information processing (#30). In sum, the above findings suggest that eye movement measures reveal the effects of background knowledge, expertise, cognitive ability and style on information processing behaviors.

Cluster 4 - Effects of instructional design on multimedia learning

The theme of Cluster 4 including eight studies emphasized the effects of instructional design on learning outcomes. Most of the studies examined the effectiveness of the uses of the multimedia learning materials which provided visual guides for learners’ attention while learning (#5, #11, #25, #27, #29). The manipulations of the design included the use of attentional guide cues (#5), pictures with color coding (#29), text with labeled pictures (#11), illustrations with changing colors for signaling (#27) and different presentation speeds of animation (#25). Regarding the other three studies, one examined the use of expert videos for surgeon skill training (#1), another investigated the impacts of different task demands for concept learning (#10), and the other probed the interaction between different types of graphical representations and uses of learning aids (i.e., uses of additional questions to organize and process information) (#32). Research topics in this cluster included conceptual development and skill learning. Both remote and mobile eye-trackers with sampling rates ranging from 30Hz to 250Hz were utilized in these studies. Almost all of the participants were undergraduate students, except for one study (#11) which involved elementary school students and one (#1) which targeted adult learners. Temporal, spatial and frequency eye-tracking measures were all shown in these studies in which total fixation duration and inter-scanning count were the most commonly used indices for uncovering learners’ visual attention distributions and transfers among different instructional elements.

The findings of the studies in this cluster mainly supported the design of visual cues or learning aids to guide students’ attention while learning with the multimedia materials. Eye-tracking analyses provided evidence that labeled or signaled pictures significantly promoted the integration of text and graphic comprehension, and therefore enhanced learning outcomes. Cueing was also demonstrated to enhance the effectiveness and efficiencies of learners’ visual searches for information relevant to the learning goals. Noticeably, the use of learning aids to support learning increased when subjects read the dynamic graphics (i.e., animations) (#32). Studies in this cluster also demonstrated the effectiveness of the reasoning task with a prediction requirement (#10) and the surgical training program using expert videos (#1). However, it is worth noting that, while manipulating the presentation speed of animation, eye movements seemed to be primarily affected by the learning content rather than by the presentation speed (#25). Eye-tracking evidence also showed the different patterns of eye movements between the groups with different learning outcomes. For example, the learners who were asked to generate predictions were found to focus more on the macroscopic video and make fewer visual transfers between the micro and the macro displays (#10). Meanwhile, longer quite-eye duration and fewer fixations were associated with better performance and more efficient visual strategies for learning surgical skills (#1).
Cluster 5 - Learning with dynamic e-platforms

The papers in Cluster 5 focused on exploring learning in dynamic digital environments in which mobile devices, interactive and collaborative software platforms, and Virtual Reality (VR) technology were employed. There are four papers in this cluster. The main participants were undergraduate and graduate students, and the sampling rate of the eye-trackers used in this cluster was either 50Hz or 60Hz. One study compared learning with desktop computers and mobile devices (#6). Another examined note-taking behaviors (with or without automatic annotation links) in the Interactive Shared Education Environment (ISEE) which supports collaborative video-based distance learning (#22). Two studies designed and employed VR learning environments (#23, #33) and discussed how familiarity and time spent with the environments may affect visual attention. Three of these papers (#6, #22, #23) dealt with learners’ conceptual development, while the other (#33) investigated perceptual learning mechanisms.

The major findings are described as follows. First of all, the use of PCs and tablets was found to be more adequate for learning than mobile phones as the use of mobile phones seemed to impose extraneous cognitive load (#6). Secondly, with the help of automatic annotation links, learners were able to attend more to the video content and annotations than to the video control operations (#22). When learners were familiar with the VR learning environments, they became more sensible to the changes of learning objects (#23). Finally, perception learning as indicated by the decrease in fixation numbers was detected when the exposure time to the VR environments changed (#33). The above-mentioned findings imply the adaptability of human learning. In sum, the studies in the cluster suggest that the above-mentioned digital devices, the online platform and VR settings might result in extraneous cognitive loads for users. In addition, familiarity or experience with the new digital devices or VR environments was found to mediate the cognitive activities engaged in the learning task.

In sum, the empirical findings of these eye movement studies provide evidence-based information that can guide the instructional design of science learning in digital environments. Based on the findings of theme cluster, we propose instructional suggestions as listed Table 1 for the design of science learning instruction in digital environments.

Instructional suggestions drawn from study findings

In Table 1, three aspects of instructional design are presented, including the basic design, dynamic environments and individual differences. The study findings of Clusters 1, 2 and 4, contributed to suggestions for the basic design. In cluster 1, it was found that learners, when reading science learning materials, attended more to verbal areas when text and graphics were presented together. When graphics contained conceptual information, the reading time of the graphics increased. The integration of verbal and visual information (indicated by the inter-scanning between text and graphics) was found to be the most significant predictor of conceptual achievement. While animations are able to depict complex phenomena or processes, they might produce extraneous cognitive load. The studies in Cluster 2 confirmed mostly the multimedia effects such as the multimedia representation effect, the coherence effect, the contiguity effect and the modality effect, but the coherence effect was not conclusive. It was shown that irrelevant illustrations in science texts moderated the beneficial effect of instructional pictures more specifically for learners with low prior knowledge. In addition, the processing time of animations predicted learning outcomes. The findings in Cluster 4 showed that visual cueing or labeled pictures enhanced learners’ attention and promoted information integration, and that learners’ visual attention and reasoning performance can be changed and improved by a prediction task. The later finding suggests that task-relevant activities given before the main learning event affect the allocation of learners’ cognitive resources.

Based on the above-mentioned findings, it is recommended that the basic design of the digital instruction for science learning should follow the design principles suggested by the theory of multimedia learning, such as placing related text and graphics near each other and employing only one type of verbal mode of information (either written or spoken). The verbal explanations should be carefully written to uncover explicitly the conceptual knowledge embedded in the graphical information. To help science learners to effectively integrate relevant information, instructional guidance encouraging back and forth scanning between different representations should be provided. In addition, effective use of visual cueing or labeled pictures will enhance learners’ attention and promote information integration. Giving learners a prediction task prior to the learning task will affect the allocation of learners’ cognitive resources and enhance learning performance. As far as the use of animations is concerned, since it has been reported that animations may cause extraneous cognitive load and require more time to be processed in depth, additional instruction or explanation of the goal and content of the animation needs to be given to guide learners’ attention.
### Table 1. Instructional suggestions drawn from the findings of eye-tracking studies

<table>
<thead>
<tr>
<th>Design aspects</th>
<th>Features</th>
<th>Major eye movement findings (Cluster#)</th>
<th>Instructional suggestions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic design</td>
<td>Verbal mode + visual mode</td>
<td>• Readers attend more to the written texts than to the graphics. (C1) • Verbal information guides visual attention (C1) • When the graphics are in conceptual, sequential or interactive form, the visual attention and processing time of the target areas increases (C1) • Transitions between different presentation modes predict concept achievement (C1) • An overview of the graphic helps to establish mental scaffolding for text reading (C1) • Visualizations were processed more when coupled with spoken texts than written texts (C2) • Irrelevant illustrations (either text or graphic) might hinder retention (C2) • Labeled or signaled pictures promote the integration of text and graphic comprehension (C4) • Cueing enhances the effectiveness and efficiencies of learners’ visual search for the relevant information (C4) • Reasoning tasks with a prediction requirement enhance performance (C4)</td>
<td>• Related text and graphics should be placed near each other (contiguity effect) • Use one verbal mode at a time (either written text or spoken text) (modality effect) • Learners rely heavily on written information to construct their conceptual models. Therefore, verbal explanations should be carefully written to explicitly uncover the conceptual knowledge embedded in the graphical information • When graphics contain conceptual information or are in sequential or interactive form, additional guidance or instruction that directs learners’ attention and processing time of such graphics is critical • Giving instructional guidance encouraging back and forth scanning between different representations will help information integration • Use labels, cues or questions to guide and encourage back and forth reading • Place learners in a learning context that requires them to reason and make predictions</td>
</tr>
<tr>
<td></td>
<td>Verbal mode + dynamic visual mode (Animation, video)</td>
<td>• Animations may result in extraneous cognitive load and require more time to process (C1) • Questions that help to organize and process information promote animation reading (C4)</td>
<td>• Allow self-paced timing of the reading of animations • Explicit explanations of the goals or contents of animations are necessary to reduce the extraneous cognitive load • Use organizer questions to help meaning construction, such as: • What is important about the statement? • How is the demonstration related to the topic to be learned? • How are different concepts/ideas related? (Note. The above questions are adopted from Ruf &amp; Ploetzner, 2014)</td>
</tr>
<tr>
<td>Dynamic environments</td>
<td>Verbal mode + visual mode (Learning with mobile devices or in a VR environment)</td>
<td>• Learning in PC environments is more efficient than on mobile devices which seem to impose extraneous cognitive loads (C5) • In the video environment, annotation links help attention allocations (C5) • Familiarity and prior experiences</td>
<td>• To enhance the effectiveness of the dynamic and interactive learning platforms, some preparation or training sections are necessary for learners to become familiar with the new devices or settings</td>
</tr>
</tbody>
</table>
with the new learning platforms (including mobile devices and the VR setting) determine the success of learning. (C5)

<table>
<thead>
<tr>
<th>Individual differences</th>
<th>Background knowledge</th>
<th>Cognitive abilities</th>
<th>Cognitive styles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Background knowledge helps to reduce processing time (C1)</td>
<td>Learners with higher spatial test performance display better working memory and employ more analytic strategies (C3)</td>
<td>Subjects identified as imagers fixate more on images, verbalizers on text, and intermediates equally on images and text (C3)</td>
</tr>
<tr>
<td></td>
<td>Learners with a strong background spend more time reading both keywords and graphics, and make more transitions between graphics and text (C3)</td>
<td>Learners with lower spatial test performance would need learning aids that can show the result of spatial manipulations in mind to support their working memory</td>
<td>Environments that match learners’ cognitive styles seem to enhance learning results (C3)</td>
</tr>
<tr>
<td></td>
<td>Decorative pictures moderate the beneficial effect for learners with lower prior knowledge (C2)</td>
<td>A training program introducing the analytic strategies for solving the spatial related problem should be helpful for learners with lower spatial performance.</td>
<td>Adopt different design approaches. For example, more visual representations should be used for imagers, while more verbal explanations may benefit verbalizers.</td>
</tr>
<tr>
<td></td>
<td>Learners with low prior knowledge need more coherent multimedia instruction (C2)</td>
<td>For learners with lower prior knowledge, avoid irrelevant illustrations and provide highly coherent instruction.</td>
<td>An instruction or guidance that can direct the attention of imagers/verbalizers to verbal/visual areas respectively, will be necessary to enhance the process of information integration</td>
</tr>
</tbody>
</table>

The dynamic digital environments are characterized by the use of mobile devices, interactive software platforms, or VR technology. These digital technologies have been developed in recent years, and therefore their effectiveness, when they are used to support multimedia learning, needs further evaluation. The studies in Cluster 5 addressed this issue. It was shown that although these digital environments have unique features such as portability, user-controlled interaction, and replication of an environment that simulates a physical presence in places in the real world, the use of these newly developed learning environments might result in extraneous cognitive load. Some studies in Cluster 5 found that experiences and familiarity with the new technologies affect cognitive activities. Accordingly, additional preparation or pre-task training programs are favored in order to enhance learners’ familiarity with the environmental settings while also reducing extraneous cognitive load.

The last aspect of the instructional suggestions is individual differences. As discussed in Cluster 3, individuals with different knowledge backgrounds, cognitive styles and spatial performances displayed different eye movement patterns, indicating a close link between visual attention and information processing strategies. The findings of these studies provide empirical information that can be used to develop adaptive instruction. For example, those who had better background knowledge were more able to allocate their visual attention to key areas in the science learning materials. Accordingly, visual guides or cues should be provided especially for science learners with lower prior knowledge so that they can allocate their attention and time more efficiently to the target areas or activities. Learners with lower spatial test performance would need learning aids that can show visually the result of spatial manipulations in mind to support their working memory. This can be achieved by the design of interactive programs allowing learners to display their mental work. In addition, these learners...
might need additional training to develop appropriate analytic strategies such as selecting, comparing, matching, orienting and so forth for solving spatial related problems. For students who are identified as imagers/verbalizers who look more at images/text, instruction or guidance that can direct their attention to verbal/visual areas, it will be necessary to enhance the process of information integration. Noticeably, although background knowledge was not the main issue addressed in Clusters 1 and 2, several studies in the two clusters addressed the issue as one of their research questions. In general, higher background knowledge less processing time. In short, the eye movement studies addressing individual differences have confirmed that attention distributions and the speed of information processing are affected by cognitive factors and prior knowledge. Accordingly, instructors should identify first these characteristics that their learners possess, and make use of the characteristics to develop adaptive instructions.

In conclusion, our analysis on the themes of the selected papers suggested that the design of e-learning instruction should consider placing related text and graphics in adjacent areas, using one verbal mode at a time, providing explicit and clear verbal explanations, and explaining carefully the goals and contents of the animations to reduce the extraneous cognitive load when animations are used. In addition, before students are placed in the dynamic learning environments, the instructor needs to provide a pre-training program allowing students to become familiar with the new environment. Finally, individual differences such as background knowledge, cognitive abilities and cognitive styles should be taken into consideration in the instructional design.

**Suggestions for future studies**

Based on the findings of this review study, we have proposed some design principles for digital learning as Table 1 shows. In addition to the suggestions for schoolteachers or curriculum instructors, there are many research issues worth of further explorations. As mentioned in the “general characteristics of the reviewed papers,” it was evident that most of these studies recruited adult students, which might limit applications of the instructional suggestions. To understand more of learners’ characteristics and development trajectories, future studies should involve younger science students, such as elementary students or even preschoolers. While eye movement measures have successfully demonstrated that differences in prior knowledge or expertise, cognitive ability or style would result in variations in information processing strategies, the studies on individual differences have not been sufficiently accumulated to be conclusive. It is thus expected that more eye-tracking studies be devoted to examining the personal factors such as cognitive ability, learning style and personal values. As discussed in several papers reviewed in the study, science learning performances were influenced by experiences and familiarity with the multimedia learning environments. Since science learning performance is also mediated by personal factors such as background knowledge, cognitive ability, and learning style as discussed in the study, another research agenda in the future should be the testing of the effectiveness of adaptive instructional designs, taking into consideration personal factors, and exploring the patterns of eye movements associated with better learning outcomes. These patterns may serve as references for providing individual feedback in the adapted learning systems.

In addition to the learner characteristics and cognitive factors, future research attention should be also placed to the learning topics and the teaching/learning materials. As mentioned in the overall results, the learning topics explored by the reviewed papers focused largely on conceptual developments. Topics such as skill learning and reasoning which have attracted educators’ attention in recent years deserve more research efforts. Among the reviewed papers, most of them used static images as their experimental materials, with only a few exceptions. Given that dynamic images or graphics are becoming a popular form of multimedia representation in science teaching and learning, more studies should be conducted to test their effects on learning. However, since tracking a dynamic AOI (such as a video or simulation) is complex and time consuming, it might need further collaboration with software or image-processing developers to analyze the eye movement data.

A review study written by Hsu et al. (2012) has shown that a growing research in the technology-based learning concerns issues regarding pedagogical design the development and evaluation of new learning systems, platforms and architectures. With the rapid development in media or visualization technologies such as augmented reality (AR) and VR technologies, dynamic learning environments have been created. By tracking eye and head movements in an immersion environment, it becomes possible to monitor the perceivers’ relation with virtual objects’ spatial characteristics (Renaud et al., 2003). Therefore, the last recommendation for the future work is the call for more eye tracking studies examining the effectiveness of learning in the new digital learning environments, in particular the dynamic types such as the mobile devices, AR and VR environments which have been developed and taken notice of only in recent years. To study the dynamic digital learning environments, the mobile types of eye trackers are preferred. Nevertheless, the software that can effectively
analyze the eye movements in moving areas of interest is still premature. The collaboration between software engineering and education researchers is thus necessary to optimize the eye tracking method.

References


<table>
<thead>
<tr>
<th>N</th>
<th>Title</th>
<th>Author</th>
<th>Journal information</th>
<th>Learning topics</th>
<th>Subjects</th>
<th>Eye movement measures</th>
<th>Performance measures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2. Number of fixations</td>
<td>2. Performance time</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3. Movement time</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2. Total fixation time</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3. Eye transitions</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2. Total fixation duration</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3. Total regression number</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4. Inter-scanning counts</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2. Average fixation duration</td>
<td>2. Concept test</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3. Percentage of total fixation duration</td>
<td>3. Responses to interview questions re problem solving strategies</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4. Number of regressions</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2. Number of gaze shifts</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seductive details and attention distraction – An eye tracker experiment</td>
<td>Rey, G. D. (2014). In Computers in Human Behavior, 32, 133-144</td>
<td>Conceptual Development</td>
<td>55 University students</td>
<td>1. Mean total fixation time</td>
<td>Knowledge tests (retention &amp; transfer)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The role of decorative pictures in learning</td>
<td>Lenzner, A., Schnotz, W., &amp; Muller, A. (2013). In Instructional Science, 41, 811-831</td>
<td>Perception</td>
<td>7th and 8th graders</td>
<td>1. Gaze duration</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

40
<table>
<thead>
<tr>
<th>Evidence From Eye Movements</th>
<th>times (look-from)</th>
<th>3. Working Memory tests</th>
<th>4. Achievement records</th>
</tr>
</thead>
<tbody>
<tr>
<td>16 Effects of different multimedia</td>
<td>N/A</td>
<td>Conceptual development</td>
<td>College students</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>Title</td>
<td>Authors</td>
<td>Journal</td>
</tr>
<tr>
<td>---</td>
<td>----------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>----------------------------------------------</td>
</tr>
<tr>
<td>Learning: Evidence from eye movements</td>
<td>1307 meiosis) age 12 yrs</td>
<td>time, mean fixation duration</td>
<td>meiosis), and retention-test scores</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>--------------------------</td>
<td>----------------------------</td>
<td>-------------------------------------</td>
</tr>
<tr>
<td>32 One click away is too far! How the presentation of cognitive learning aids influences their use in multimedia learning environments</td>
<td>Ruf, T., &amp; Ploetzner, R. (2014)</td>
<td>Computers in Human behavior, 229-239</td>
<td>Conceptual Development University students 1. Fixation counts 2. Fixation sequence saccade (alternating between reading and in the support and content areas) 1. Posttest on factual knowledge 2. Transfer test</td>
</tr>
<tr>
<td>load theory</td>
<td>velocity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. Pupil</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>diameter</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
High School Students’ Views on the PBL Activities Supported via Flipped Classroom and LEGO Practices

Baris Cukurbasi* and Mubin Kiyici

*1Department of Computer and Instructional Technologies Education, Faculty of Education, Bartın University, Bartın, Turkey / 2Department of Computer and Instructional Technologies Education, Faculty of Education, Sakarya University, Sakarya, Turkey // bariscukurbasi@gmail.com // mkiyici@gmail.com

1Corresponding author

(Submitted October 26, 2016; Revised January 11, 2017; Accepted January 29, 2017)

ABSTRACT

The purpose of this study was to investigate the high school students’ views on instructions based on Flipped Classroom Model (FC) and LEGO applications. The case study, which is one of the qualitative research methods, was used within the scope of the study, the duration of which was 7 weeks. In order to choose the research group of the study, one of the purposeful sampling methods, criterion sampling, was used. In this context, the study was conducted with 35 10th grade high school students from two different classes. EGA (Experimental Group A) consisted of 18 students while EGB (Experimental Group B) consisted of 17 students. Within the scope of the study, algorithm instruction with FC and LEGO applications was conducted on EGA and LEGO applications and algorithm instruction in classroom were conducted on EGB. In accordance with the study, processes of focus group discussions and gathering data through observation were carried out. For the analysis of the data that was gathered, content analysis was performed and NVivo 8 was used. In the end of the study, the students were observed to be having great prejudices and reported negative views in the beginning of the study. Upon the initiation of the study, the prejudices and negative opinions were observed to be replaced by positive opinions. Students mentioned a great deal of educational benefits of the study. They stated that with the study, their motivation and interest in the lesson increased thanks to the implementations. Additionally in the group work, students said that they cooperated, exchanged ideas, shared tasks, took responsibility and socialized with their friends. About FC, students gave positive feedback on the communication opportunities of the atmosphere and they stated that they could repeatedly go beyond the class hours with FC, hence came prepared and saved time. Also students stated that FC atmosphere improved opportunities for teacher-student communication.

Keywords

Flipped classroom, LEGO, Robotic, Algorithm instruction, Problem-based learning

Introduction

It is seen that new technological products and developments that came to light each day are involved in the process of education (Uğur Erdoğan & Cagiltay, 2013). Additionally, it is seen that there is a rapid change in the learning and teaching methods. As a result, with the help of ICTs education processes are made easier and permanent learning is enabled (İsmahan, 2011). These developments also bring more varying learner characteristics compared to past. The era we are in is called the digital era (Spranger, 2010). Prensky (2001) called the individuals who were born in digital era digital natives. These individuals who were born in the digital era always interact with ICTs from the day they were born. Digital native students always carry phones, tablets and other ICT tools with them, trust technology that is present and live with these technologies (Oh & Reeves, 2014). In the classroom, these students develop new roles such as researcher, technology user and expert, reasoning and inference, using information that is learned in real life and self-teaching (Prensky, 2010). Hence it is thought that ICT use will be important for digital native students in acquiring and processing the information and thus, making teaching more efficient. With the inclusion of ICTs in education, the concept that is defined as using necessary information and skills to control education and structuring learning and teaching processes has emerged (Alkan, 2011). In the study conducted by Woolf (2010), it is seen that as a result of education technologies and learning environments, there will be considerable changes in education. In the Horizon Report that is regularly published by NMC, the fast effect of the development in ICTs on education processes can be seen (NMC History, 2014). MOOC, tablets, cloud computing, mobile learning and Flipped Classroom (FC) are some of the subjects that are stated in the report as planned to be included in education in five years (Johnson, Adams Becker, Estrada & Freeman, 2014a; 2014b). In another study, learner analytics, cloud computing, mobile applications and MOOC are examined in the context of teaching technologies as future implementation, inclination and approaches (Baran, 2013). In addition to these, it is among the subjects that draw the robotics which interact with the users physically or verbally and are used for moving objects in virtual or real world are becoming increasingly popular (Johnson et al., 2016). Robotic objects attract many students and today robots that are bought reach students at K-12 level with their feature of programmability for various tasks (Prensky,
Besides, it is emphasized that educational robotic applications boost students’ motivation and are useful teaching materials (Ortiz, 2015). LEGO Mindstorms products are among the robotic products that are widely used in educational processes around the world.

FC model which is popular today and becoming more popular, enables the students to perform real-life applications more actively in order to understand the subjects profoundly with project based or problem based learning (PBL) applications within the limited class hour (Johnson et al., 2014a). Teacher adapts a student-centered approach as the primary teaching method instead of directly teaching (SAMS & Bergmann, 2013). In other words, with FC, cooperative activities take the place of the normal classes (Chen, Wang, Kinshuk, & Chen, 2014). With the applications related to FC, firstly the subject that needs to be taught in classrooms is converted to teaching materials that are usually videos by the teacher and then published via the Internet technologies. At the same time teacher plans the teaching activity that will be carried out in the class. Accordingly, the planned activity should be a student-centered one. Student gains information and listens to the subject that will be taught by watching/examining the materials that are published outside the classroom. Then, with the teaching activity that teacher planned, classroom environment in which the student is more active is created and with that, it is aimed that the student will learn the subject that s/he studies outside the class in a better way. Just like the developments in education technologies, there have been new approaches to cooperative learning and application described as student-centered in learning-teaching approaches in the last two decades (Cetin, 2013). PBL is one of the recent and popular approaches. PBL basically aims to face the students with situations similar to the ones they might face in their future careers and help them learn to solve these real-life problems (Erdem Gurlen, 2011). The most important point in PBL applications is to eliminate the problems that occur in the process of including students in the learning.

In today’s learning environments, generally constructivist learning approach is the basis. In the process of students interacting in FC environment, performing activities according to the situation that is determined with LEGO applications, it is thought that constructivist learning approach can be supported and believed that learning and teaching activities can be done more effectively by supporting this process with PBL applications. Especially because nowadays students are digital natives, it is believed that Facebook that will be used as FC environment, LEGO applications, FC Model and PBL approach that will be applied will be more effective in using digital native characteristics. Furthermore, it is of importance to minimize problems that emerge in respect of integrating students in the process of learning. At this point, the effects of the use of LEGO applications and problem-based learning applications along with the FC environment are to be identified with a view to increase motivation of students and to ensure them to take part at the center of the learning process. Nevertheless, there has been no study to analyze the effects of problem-based teaching applications on learning and teaching processes when they are realized through use of a combination of robotic applications such as LEGO and environments such as FC, among today's popular technologies, which, in addition to other factors, has necessitated a study on this subject. So, the study aims to examine the opinions of high school students on problem-based teaching activities supported with FC Model and LEGO applications.

With this purpose, the study seeks answers to the following questions, (1) What are the students’ views on problem-based LEGO applications? and (2) What are the views of students on PBL activities supported via FC?

**Theoretical background**

**LEGO applications during education**

LEGO education department stated that children should be supported to become systematically creative, active and cooperative learners in the manifesto they published (LEGO Education, 2010). They developed LEGO Mindstorms products to carry out educational LEGO applications. Learning environments that are created in LEGO applications include constructivist approach (Cayir, 2010; Danahy et al., 2013). According to constructivist approach learning occurs by learners’ active participation in the process of making meaning and is described as an active process which is developed by the experience of learners (Tufekci Aslim, 2013). In this context, LEGO applications as education tools enables students to learn actively, create a constructive environment, physical objects and find ways to make meaning of abstract concepts (Chambers, Carbonaro, & Murray, 2008).

Some of the benefits of LEGO applications for educational process are as follows:

- LEGO enables each student to create different solutions to the same problem, that is, help them improve their problem-solving skills (Cavas et al., 2012; Danahy et al., 2013; Lin et al., 2009).
As a sharable structure, it makes inquisitive, cooperative and constructive learning easier (Chambers et al., 2008; Koc Senol, 2012; Ozdogru, 2013).

It is one of the most suitable tools for improving learners’ creative thinking (Lin et al., 2009).

It allows students to improve their group work and communication skills (Aufderheide, Krybus & Witkowski, 2012).

It enables students to be more independent and confident learners (Church et al., 2010).

Students get the chance to get deeper information, think limitlessly and practice their ideas (Cayir, 2010).

Students from every age become the active leaders of their own learning by making robots (Danahy et al., 2013).

Students’ motivation increases (Aufderheide et al., 2012).

**Flipped classroom (FC) model**

Learning environments are moved to more dynamic and social areas in order for students to be able to work in groups to solve a problem or discuss (Johnson et al., 2014b). FC is one of the models that enable learning environments to such areas. The purpose of the model that makes the applications of student-centered approaches easier is to create active environments in the classroom (Brown, 2012). Within FC, class materials are used outside the classroom and class hour is used for student-centered activities to make the subject clearer and reinforce learning (Mason, Schuman, & Cook, 2013).

In FC environment, generally videos are prepared as pre-class learning materials (Long, Logan, & Wough, 2016). In these videos, the content is organized in such way that students can study before the class. The videos can be prepared with interactive factors (question-answer, attractive items, etc.) as well as in the form of plain narration (Temizyurek & Unlu, 2015). It is suggested that videos are made interactive with Web 2.0 tools (Basal, 2015). Besides the videos in FC environment, Bergmann and Sams (2012) pointed out that materials such as books and documents could be added. In this context, it is believed that teacher has important roles of preparing materials for FC environment and classroom activities.

Teacher roles in FC environment differ from those in traditional classes. Here, teacher gives the lesson content to the students via videos and helps them understand the subject that students study through the videos with the help of activities that include cooperation and interaction (Mok, 2014). So, FC provides students with individual or cooperative problem-solving activities in classroom on the subject that students study outside the classroom (Gencer et al., 2014). FC applications throughout the world are observed by many researchers and applications for K-12 level are carried out (Chen et al., 2014). As a result of these applications with FC Model, teacher is no longer the one who presents the information and students are the ones who manage the learning process (Brown, 2012). Also because this model uses technology to which today’s digital native students are used and class hours are saved for activities, it is thought that FC Model is suitable for digital native learners.

FC Model improves learning outcomes, support active learning and high level thinking (Baepler, Walker, & Driessen, 2014). At the same time FC supports technology use for teaching outside the school (Herreid & Schiller, 2013), gives responsibility of gaining knowledge to the learners (Butzler, 2016). FC application also boosts motivation (Strayer, 2012; Turan, 2015), and improves student learning performance (Hung, 2015). In addition to these, with FC, students find the opportunity to learn individually and in this context they can adjust their own studying time flexibly (O’Flaherty & Philips, 2015).

**Method**

Case study which is one of the qualitative research methods was used in the study.

**Participants**

To choose the research group, criterion sampling which is one of the purposeful sampling methods was used. In the context of criterion sampling, completely voluntary students of vocational high school that suited the criteria of having more than one 10th grade classes of Computer Technologies Department (CTD) and at least two classes that take Basics of Programming class from the same teacher took part in the study. So, the study was conducted with the Experimental Group-A (EGA) that consisted of 18 students and Experimental Group-B (EGB) that consisted of 17 students, 35 students in total from two different classes. 9th grade students having
been educated at different classes are selected to CTD as regards their arithmetic scores. 10th grade students starting to be educated at CTD don’t know each other at the beginning of the semester.

Data collection tools

Focus group interview form

A 10-item semi-structured focus group interview form was created by the researchers. After 7 weeks of work and observation, the form was reorganized (see Appendix 1). Lastly, an 11-item interview form for EGA and a 9-item interview form for EGB were created.

Observation

The works were recorded on camera and observations were made by the researchers. The camera recordings were watched and observation records were edited and then observation data was produced.

Implementation area

Implementation is made in a computer lab in the high school (see Figure 1).

![Figure 1. Implementation area](image)

Implementation process

The study was conducted in the first semester of 2015/2016 school year. In the 7-week-long study, a work on teaching algorithm with FC and LEGO applications to high school students was done. There was interaction with the students except for the first week that was spent with creating the study groups. For the students in EGA to be able to work according to FC, a private group (FCG) was created on Facebook (see Figure 2).

![Figure 2. FCG](image)

FCG included all the helpful educational content as videos in teaching algorithm and flow charts, as well as LEGO Mindstorms EV3 set that students would be using as class content and EV3 programming. These videos were not uploaded on FCG all at once; rather, they were uploaded one by one. 22 videos between 4-20 minutes
were uploaded on FCG. On the study day, all the time that was saved was spent with applications. With EGB, all work was done in the classroom face-to-face with the students. Schematic presentation of the pre-lesson, in-lesson and post-lesson processes with the study groups are given in (see Figure 3).

**Figure 3. Process of teaching**

Students were asked to design a robot and program those robots, and then PBL themes (PBLT) were created by the researcher for the students to use LEGO applications in teaching algorithm. The main subject of the PBLTs was “Let’s Create Solutions for the Aging Population” (see Figure 4).

**Figure 4. PBLT images**

With PBLT, some of the problems that the elderly may come across were pointed out in the themes. Students were asked to create solutions to these problems by improving and programming their robots. There were four different problems on each PBLT area. With the PBLT document and narration which was given to EGA on FCG and EGB in the classroom, the situation and rules that they had to follow to solve the problem were explained to the students. Before starting to solve the problems with the robots, students were asked to prepare the algorithm and flow chart of the solution. At this point, it was explained that the students could prepare the algorithms and flow charts in cooperation with a part of their friends or all of them as well as individually.

At the beginning of the study, the teacher of the class was briefed about the study. Then considering the timetable, research groups were chosen with the teacher. Taking account of the sizes of the classes, each research group was divided into two team among themselves with a random nickname (EGA=AG1+AG2, EGB=BG1+BG2). Each team was provided with a EV3 set (see Figure 5).
During the 1st, 2nd, 3rd, 4th weeks, students were instructed on the EV3 set (see Figure 6). In these weeks, only EV3 use, robot design and programming are taught.
Students designed their robots and learned how to program them. In the 5th, 6th, 7th weeks, students worked on PBLT area. In this context, students prepared algorithm and flow chart at the stage of solving the problem on PBLT area and programmed their robots. After the implementation process, the teacher carried out lessons in the first half of the students’ normal class hour. In the other half of the lesson, LEGO applications were performed. In the week that algorithm and flow chart were taught, students programmed their robots and completed the work by solving the problems on PBLT (see Figure 7).

![Figure 7. 5-7 Weeks of works](image)

At the end of the study, a focus group interview was done with the groups.

**Data collection**

Separate focus group interviews were done with the groups in EGA and EGB. During the whole process of the study, all the work was done and the students were observed by the researcher. At the same time, the work was recorded on camera. After each week, notes were taken on the works by the researcher. The points that went unobserved and all work in general were examined on the camera record.
Validity and reliability of the study

The researcher interacted with the students in person in times of the work and on the Internet with EGA outside the work time. Students answered data collection tools more sincerely and easily. Also, the data was collected through observation and interview to be able to examine the students’ emotions, thoughts and behaviors towards the study. After the process of data collection, the findings and results that were gathered in the process of data analysis will be compared with each other and the relationship between them will be revealed. In the study, criterion sampling from purposeful sampling methods was used. During the focus groups interviews, the questions were asked in the same order to each participant. While analyzing the data, no comment was added by the researcher. Besides, consistency of the analysis results was compared by making two experts analyze the data. In this context, the reliability of the study was determined by using reconciliation percentage formula by Miles and Huberman (1994).

\[
Compromise\ Percentage\ (P) = \frac{512}{512 + 42} \times 100 = 92.42\%
\]

Data analysis

NVivo8 was used to analyze the data gathered from the study. Focus group data were examined through content analysis. In the analysis process, the data were divided into three groups of pre, while and post implementation. The views on FC were divided into a different group. Besides the researcher, the data were analyzed by two academicians and the results were compared. Then the focus group interviews were transferred into electronic media. After that, it was analyzed through NVivo8. In this context, the data were coded and assigned to lists and then put into themes. At the end of the focus group interview analysis, the data were supported with observation results and meaningful results were attempted to create.

Findings

This section mentions the results that were reached at the end of data analysis. The statements of participants were given beside the findings. In the thematic displays that were shown in the findings section are frequency of values and first value is shown as EGA, second value as EGB (EGA-EGB).

Findings on the students’ views on LEGO applications

Findings on pre-implementation

The thematic display of students’ views on pre-implementation process is given below (see Figure 8).

![Figure 8. Thematic display of pre-implementation results](image-url)
When Figure 8 is examined, it can be seen the students were mostly prejudiced against the study. The thought of giving up the work, unwillingness, getting bored were some examples of students’ statements. Besides these, they used statements that included fear, anxiety and unwillingness. The statements quoted from students are as follows:

“At first I thought I was going to fail at such a class, with the robot application. I mean I would be negatively affected. As I continued to do it, I had fun” (AGS15).

“When you first came here and we watched, I thought we could not do it. I thought ‘Are we really going to create a robot?’ I thought I would have so much difficulty at first” (BGS14).

It is seen besides the statements of, there are also ones that include curiosity, willingness and excitement. In this context, students stated they were interested in the study, curious about it and thought it would be exciting. The positive statements quoted from the students are as follows:

“At first I was intimidated when I saw the pieces. We did not want to come to the class. Then we put the pieces together and something good came out” (AGS14).

“Honestly I was really excited when I first saw the robot” (BGS16).

It is seen that the students were shy in pre-implementation process. This may be due to the fact that the researcher and students met for the first time, newly starting the lesson and did not know each other.

Findings on while-implementations

The thematic display of students’ views on while-implementation is given below (see Figure 9).

![Thematic display of while-implementation](image)

When Figure 9 is examined, the views such as that the process was informative, students learned by trial-and-error methods and that they had a feeling of success when they solved problems on LEGO applications were stated. Besides this, the students stated that their abilities to make plans and manage the time were improved at the end of the study. The statements quoted by students on educational benefits are as follows:

“We got the best of it. We learned how to plan and solve problems. So it was good” (AGS13).
“We did not know much about algorithm and flow chart. Then we did things like these and it was good with robots. We test and learn better.” (BGS4).

Students stated that when they work in teams, they worked cooperatively with their teammates. Helped each other and exchanged ideas between teams. About task-sharing, students’ statements are as follows:

“Working in groups means producing different ideas in a short amount of time. When one of us could not do something, the other helped him do it” (AG15).

“If we tried to do it alone, we would not be able to continue. When we do it in groups, everyone produces an idea. So group work is more advantageous” (BGS17).

On the work done, the students said that taking responsibility was important and that everyone worked according to it. They also stated that LEGO applications were functional and they could be used not only to solve problems on the area but also to solve other possible problems in daily life. Besides, students said that with the study, they got to know their classmates better and interacted with them. About the daily life theme, students’ statements are as follows:

“We got to know our classmates. We understood what we can accomplish with them” (AGSI1).

“There are disabled people. If robots are made with the purpose of helping people, they can help disabled people or women, in the kitchen” (BGS3).

Students stated that implementation process was fun and increased their interest in the lesson. The studies drew their attention and the study in general boosted their motivation. They also stated that they worked with excitement. About the affective features theme, students’ quoted statements are as follows:

“...Then it became more fun when we started to create the robot” (AGS8).

“If you did not start this with this class, it would not be fun at all. I did not attend the classes. We did not use to have fun. It just passed like that” (BGS5).

Students said that they could revise their works, self-assess, gauge the weak/strong points of themselves and their robots. About assessment skills, statements of students quoted as follows:

“When we first did it, the robot was very long. Then we made another plan and this time it was shorter and with stronger pieces” (AGS15).

“It was simple in appearance. The functions are good. But I think the appearance was simple” (BGS6).

When the observation data on implementation process was examined;

- At the stage of deciding what kind of robot they would make, it was seen that every student expressed an idea, discussed and exchanged ideas. It was seen that every student made research and analysis at this stage.
- It was seen that on the robot designs PBL applications, all the students contributed and helped each other.
- It was seen the students had fun during the implementation and that they did not leave even on the breaks. So it can be interpreted as the high motivation and interest towards lesson.
- While testing the robots, it was observed other teams also watched it, exchanged ideas and revised their own robots.
- When EGA and EGB were compared, although the algorithms and flow charts the students prepared prior to PBLT did not clearly reflect the solution, it was seen that it helped them follow a path on the solution through observation and on student feedback it was seen that after solving the problem with robot programming, the algorithm and flow charts reflect the complete solution of the problem.

Findings on post-implementation

Thematic display of the students’ views on post-implementation is given below (see Figure 10).
When Figure 10 is examined, students stated they mostly learned algorithm and flow charts subject and supported it with repetitions. Besides, students stated they felt successful when they solved problems with LEGO applications. The students’ quotes on educational benefit are as follows:

“They said programming classes were difficult. This year we understood it better with robot.” (AGS4).
“The most beneficial part for me is that I did not use to know how to do algorithm before you came. It seemed complicated to me. When you came, it got better with the robot” (BGS11).

Students stated that they socialized with their classmates during the study. So, this situation can be interpreted to have improved the students’ communication between themselves as a team. Besides, the students stated their opinions on future works. The students’ statements quoted on daily life theme are as follows:

“Not that robots we built but in the future, bigger ones might help in carrying cargos or in a company” (AGS15).
“I think we benefitted too. Besides, it was a good memory for us. So it was not just a class, but at the same time something that I discussed with my friends” (BGS6).

Lastly, the students stated that the study increased their self-confidence and interest in the lesson. On affective features theme, the students’ statements were quoted below:

“We applied algorithm not by directly listening to the teacher, but by having fun and making it” (AGS15).
“It was fun. Before the robot it was boring” (BGS8).

Findings on the students’ views on FC applications

Thematic display of students’ views on FC applications is given below (see Figure 11).
When Figure 11 is examined, students stated that they contacted their teacher via FC and asked questions about their works. Students said they could watch the videos repeatedly, revise the subjects and exchange ideas with their classmates. Also, the students said watching the videos before the class saved time and helped them come to class prepared. Quotes on EGA's positive opinions on FC environment are as follows:

"Without Facebook group you would have to teach it here. We would spend more time in vain” (AGS14).

"For example, we can prepare for the content by watching the videos. And there is always someone who cannot see while working so they can watch the video and see how it is done” (AGS15).

As negative views on FC, students in EGA stated they could not spare time to watch the videos due to their work/social lives. Besides, students said that they (on the first weeks of the study) had the prejudice of thinking that work in the classroom could be done without watching the videos and did not do necessary work on FC environment. Also, some students said it would be better to apply what was watched on FC environment immediately. On the theme of negative opinions on FC environment, the quotes of the students are as follows:

"There was a connection problem while watching the videos. I rewound and watched again” (AGS7).

"Class environment is more efficient. We work with our friends. On Facebook, some watch the videos and some do not” (AGS11).

Discussion

It was seen that LEGO applications had a positive effect on motivation and motivation level of students at the end of the study were higher than the beginning. The other studies concluded LEGO applications boosted students’ motivation (Ortiz, 2015; Blikstein, 2013). Besides, it is thought FC applications done with EGA also had an effect on students’ motivation. This result resembles the other studies in this field (Basal, 2015; Hsieh, Wu, & Marek, 2016). It was stated by the students that LEGO applications that were done in groups cooperatively improved the communication between the students and their teacher. At the beginning of the study, none of the students knew each other. So students had an opportunity to get to know each other with group works. Similarly, it was seen that students’ communication process with the researcher whose role was the same with the students increased. When views of EGA and EGB were compared, it was seen that EGB stated more views on communication than EGA. However, when questions were asked to EGA about FC environment, most of them gave positive feedback on teacher-student relationship. This is thought to stem from the fact that EGA had the opportunity to communicate with the teacher outside the class via FC environment and ask questions without having to wait for the class like EGB. FC enables effective teaching and makes way for student-student, teacher-student interaction (Mehring, 2014). Students in EGA said that they could contact their teacher outside
the classroom, work with their friends and exchange ideas. Similar to the studies in the field, it was seen that the fact that students could contact with their friends and teacher pleased them (Turan, 2015). Peterson (2016) stated in his study that the fact that teacher answers students’ questions effectively in FC environment was among positive feedback. LEGO applications were carried out in groups cooperatively. LEGO applications generally help student work in groups and learn cooperative learning (Auferheide et al., 2012). Besides it was stated that the fact that students worked cooperatively was among the benefits of FC (Gross, 2014). Strayer (2012) stated that with FC, students worked cooperatively and more easily.

There were negative views on FC. Though, these negative views disappeared over time. Similar studies showed that students had negative opinions when they first came across FC and then these views changed (Goru Dogan, 2015; Turan, 2015). It can be interpreted as students in Turkey not being familiar with FC and such environments. So, the process should be made easier for students (Chen et al., 2014). Besides, students should be made familiar with using FC by regarding cultural differences and learning styles (Goru Dogan, 2015). Within the scope of this study, students in EGA were told they would get extra performance grades if they came prepared and that week the majority of the students came prepared on FC. Kim, Kim, Khera, and Getman (2014) suggested students should be presented with encouraging factors to come prepared for the class. Similarly, Turan (2015) emphasized reinforcements should be used to encourage students. So in these studies, students may be told they are going to be graded or other kinds of encouragement may be given.

It was said to be positive by the students that they could watch the videos on FC repeatedly and whenever they wanted. The other studies reached similar results (Touchton, 2015). Students also stated that they could not find time to watch them because they were too long or took time. A similar result was found in another study (Turan, 2015). It is thought that FC technology plays an important role in fixing these problems. Because instead of Facebook that was used in this study as FC environment, another one that allows videos to be uploaded all at once or in one frame and allow students to choose could be used. The study suggests that an FC environment appropriate for allowing interaction and controlling the students should be used (Filiz & Kurt, 2015).

Conclusion

- It was seen that the students were generally prejudiced against the study. However these negative opinions gave place to positive ones.
- It was concluded that there was no negative feedback from students in the implication process. They pointed out the importance of the study on educational benefits, group work and improving daily life skills. Besides. With LEGO applications, it was seen that it helped students learn through trial-and-error, decide if they succeed and take responsibility in work. Also, the study reached the conclusion that the application boosted students’ motivation, attracted them and was fun.
- At the end of the application, it was concluded that students thought they succeeded with LEGO applications, they learned better, socialized, and improved their problem-solving skills and their interest in the lesson increased.
- It was concluded that LEGO applications helped students in creating algorithm and flow chart.
- The conclusion was that FC enabled students to revise, watch videos repeatedly, and come prepared, save time no subjects that would be taught in class and work on the project instead and that this generally contributed to the lesson. Having said that, it was also seen that the students came home tired due to full-time school and having sports after school and could not do enough work on FC. However this problem disappeared when students were provided with enough reinforcements.

Acknowledgements

This research was drived from doctoral dissertation titled Examine the Effect of the Problem Based Learning Activities Supported via Flipped Classroom And LEGO-LOGO Practices On The High School Students’ Success and Their Motivation. This work was supported by the Sakarya University Scientific Research Projects Committee (Project number: 2015-70-02-001).

References


An example of the use of technology as a material in language teaching: “Flipped Classroom”


LEGO Education. (2010). A System for learning. Retrieved from https://doclegend.com/queue/lego-education-manifesto-final_59df20c2d64ab2e01de2f33c_pdf?queue_id=5a571c0bd64ab2d9ec36f736


Ozdogru, E. (2013). Fiziksel olaylar öğrenme alanı için LEGO program tabanlı fen ve teknoloji eğitiminin öğrencilerin akademik başarısına, bilimsel süreç becerilerine ve fen ve teknoloji dersine yönelik tutumlarına etkisi [The Effect of Lego programme based science and technology education on the students academic achievement, science process skills and their attitudes toward science and technology course for physical facts learning field] (Unpublished master dissertation). Dokuz Eylul University, Izmir, Turkey.


Appendix 1

Interview questions example

EGA interview form
1- Explain your thoughts when you have learned designing robotic applications with LEGO products?
3- Tell about your experiences on the process in which you have started to learn LEGO applications?
5- What are your thoughts on Flipped Classroom?
7- What are your thoughts on the robot created as a group?
9- What do you think about the effect of robotic applications on learning algorithm and flow charts?

EGB interview form
2- How do your initial thoughts on LEGO applications change when you have examined the LEGO Mindstorms EV3 set for the first time?
4- What are your thoughts on preparing studies about robotic applications, algorithm and flow charts?
6- What did you get after the robot applications? What do you think about the contribution of the robotic applications on learning algorithm and flow charts?
8- What do you think about the contributions of these studies to you in the future?
Evolving Learning Paradigms: Re-Setting Baselines and Collection Methods of Information and Communication Technology in Education Statistics

David Gibson1*, Tania Broadley2, Jill Downie1 and Peter Wallet3

1 Curtin University, Australia // 2 Queensland University of Technology, Australia // 3 UNESCO, Paris, France // david.c.gibson@curtin.edu.au // tania.broadley@qut.edu.au // j.downie@curtin.edu.au // p.wallet@unesco.org

*Corresponding author

(Submitted November 22, 2016; Revised January 4, 2017; Accepted January 9, 2017)

Abstract

The UNESCO Institute for Statistics (UIS) has been measuring ICT in education since 2009, but with such rapid change in technology and its use in education, it is important now to revise the collection mechanisms to focus on how technology is being used to enhance learning and teaching. Sustainable development goal (SDG) 4, for example, moves beyond measures of access and increasingly focuses on the sustainability of education including issues of educational quality and student outcomes. A reassessment of how ICT in education is measured to support the attainment of the SDGs by 2030 is thus a timely endeavour. The paper discusses four aspects: (1) evolving mission, methods and core principles of ICT in learning and teaching; (2) nature of ICT in education in accelerating the emergence of new learner-centred pedagogies; (3) types of learning activities associated with the use of ICT including those for leaders, teachers and students; and (4) usage and deployment patterns. This paper proposes extensions and adoptions to the existing UIS data collection instrument to enrich its capacity for understanding how ICT is being used in learning and teaching.

Keywords

Evolving learning paradigms, ICT, Learner-centred pedagogies, Teaching and learning indicators

Introduction

Since 2000, there has been enormous progress towards achieving global education targets as part of the Millennium Development Goals and Education for All targets including making progress towards universal primary education, greater numbers of girls in school, falling numbers of out-of-school children, and a dramatic increase in literacy rates. In the new Sustainable Development Goals (SDGs), Goal 4 reaffirms the belief that education is one of the most powerful and proven transformative vehicles for sustainable development. This goal ensures that all girls and boys complete free primary and secondary schooling by 2030. It also aims to provide equal access to affordable vocational training, to eliminate gender and wealth disparities, and achieve universal access to a quality higher education. ICT in education and its significant role was also recognized at the International Conference on ICT and Post-2015 Education where Article 10 states: “Successful integration of ICT into teaching and learning requires rethinking the role of teachers and reforming their preparation and professional development. It calls for promoting a culture of quality in all its aspects: staff support, student support, curricula design, course design, course delivery, strategic planning and development” (UNESCO, n.d.).

While information on the physical infrastructure of ICT in primary and secondary education has been collected and studied during the last decade (Scimicca et al., 2009; UIS, 2009), understanding the impact of ICT on pedagogical processes and educational outcomes is paramount for the next stage of policy and implementation. However, measuring the usage patterns of information and communication technology (ICT) in teaching and learning is a challenge for any country. Two challenges arise in particular: the local relevance of the information and the focus of that information on improving the experience of teaching and learning with educational technology. For example, the problematic validity of data sources has been noted by a World Bank report from the Systems Approach for Better Education Results (Trucano, 2012), when a collection activity “comes from data sources outside of the education sector itself and does not appear to be gathered according to common methodologies and definitions” (p. 106). Large-scale efforts such as the International Computer Information Literacy Study (ICILS) provide useful information for comparable systems views but often leave local educators without actionable information. The proposed indicators attempt to address these two issues of validity and relevance to local schools (Fraillon, Ainley, Schultz, Friedman, & Gebhardt, 2013).

Studying the introduction of new technology in education some thirty years ago, Plomp and Akker (1988) noted the lack of research-based knowledge about usage patterns such as the method, frequency and intensity that teachers were using computers to enhance learning as well as the resulting impact on educational practice including the school curricula. This finding still resonates today with current research that often fails to provide hard evidence of the impact of ICT on teaching and learning. While government policies often focus on...
implementation of infrastructure, access to technology, and teacher professional development, there is still not much known about the implementation and impacts of ICT in learning and teaching practices in many parts of the world. A 2011 report on Latin America and the Caribbean, for example, noted significant gaps and a clear lack of records concerning educational technology implementation (Hinostrosa & Labbe, 2011). Uma and Arulchelvan (2012) confirm in a study in India that there is a dearth of information about student use of technologies; while students are using ICT extensively outside of school, they engage in limited usage for formal in-school learning purposes and still rely primarily on traditional classroom teaching and textbooks for academic progress. In Africa and the Middle East, Isaacs (2012) noted that for many years, the focus of investments was on making successive waves of new technologies work in resource-poor education environment; an emphasis that tended toward a techno-centric approach to ICT in education.

To a large extent schools in some countries are still teaching a series of disconnected subjects using methods and structures that compromise motivation, engagement and deeper learning for superficial coverage of material expected on tests (Baker, 2007; Black, 2000; Koch & DeLuca, 2012). In contrast, the evolving paradigms of learning discussed in this paper, which have come from massive shifts caused by ICT practices in the global economy and culture, are beginning to change how people learn in informal and lifelong learning contexts and are emerging as best practices in formal educational systems. How quickly, how extensive and to what ends they are emerging depend on local education policy, funded commitments, access, schooling culture, teacher capacity and professional development, the adoption of lessons from learning sciences and sufficient reflective experience in the integration of ICT in teaching and learning.

The interaction between the teacher and the learner is also being transformed and expanded by technology-enabled interactions and capabilities. The implications for teachers in terms of roles, pedagogy and approaches is well documented in the UNESCO ICT Competency Framework for Teachers (UNESCO, 2008). While some developed countries are advancing with ICT through the use of emerging tools and practices such as open educational resources (OER), social networking and a flattened world of information and communication technologies, other countries are progressing more slowly due to a variety of financial and policy constraints.

Beginning to address the gaps in knowledge about the use and impact of ICT in education and indicative of the evolving paradigms of learning with technology, the ICILS 2013 collected information from students, teachers, school technology coordinators and school principals. The constructs of that instrument point to a shift away from collecting information about infrastructure and access toward the use of technology to achieve educational benefits, especially preparing students to participate fully in the digital age. Computer and information literacy in this context is defined as “an individual’s ability to use computers to investigate, create, and communicate in order to participate effectively at home, at school, in the workplace, and in society” (Fraillon, Schulz, & Ainley, 2013b, p. 17). Our proposed changes to international indicators follow as well as contribute new ideas to this trend.

The foundation for this paper stems from a vision to build capacity in global youth to support the goals of a knowledge economy and to achieve increased levels of digital literacy across society.

VISION: A student who graduates from secondary education is ready for lifelong learning and using ICT for personal and professional productivity. Such a student is ready to use ICT to contribute to society, start a business, succeed in tertiary education, and to start or work for a local or global company.

The paper provides recommendations for new global indicators that go beyond basic infrastructure to focus on evolving learning paradigms for ICT in education. The primary focus is on what teachers and students are doing with ICT to enhance learning country-by-country, school-by-school, and classroom-by-classroom. The proposed dimensions for the indicators form a framework for baseline and annual progress monitoring using core ideas of what it means to be a successful global citizen with a high level of ICT literacy. Furthermore, in order to represent a holistic dataset of the global education system, this paper highlights the growing need for inclusion of data at the post-secondary level, thereby including the higher (tertiary) education sector in all aspects of the new indicators, including the need for disaggregated data by gender, location and socio-economic status to shed light on internal disparities within countries to support policymaking that aims to ensure there is balance in providing access to ICT and educational opportunities in order to maximize all human potential.
Previous data collections on how ICT is used in teaching and learning

A recent study, which summarizes 25 previous meta-analyses based on more than 40 years of research, concluded that computer use in the classroom does have an overall positive effect on achievement (Tamim et al., 2011). International and national student assessments have attempted to establish some relationships between usage of computers and student outcomes. For example, the International survey data from the Programme for International Student Assessment (PISA) of 15-year-old students in Organization for Economic Co-operation and Development (OECD) countries sheds some light on the relationship between intensity or time using computers and performance in mathematics and science. For many countries in reading, mathematics and science, performance scores are generally curvilinear, whereby students who are not using computers during class as well as those using computers the most score higher than those using computers in moderation (Organization for Economic Co-operation and Development, 2015).

Caution should be taken when interpreting the OECD data however, since: (i) lower-performing students are often assigned disproportionate time using computers for remedial purposes; (ii) it is difficult to isolate the influence of home computer use on school performance; and (iii) not all countries have successfully integrated ICT in pedagogically meaningful ways. One interpretation of the failure of ICT usage to be positively related with high student achievement is that building deep, conceptual understanding and higher-order thinking requires collaborative social interactions with peers and experts, often missing in ICT implementation (Griffin, McGaw, & Care, 2012; Webb & Gibson, 2015), where inappropriate use of technology can distract from this valuable human engagement. Another interpretation is that we have not yet become good enough at the kind of pedagogies that make the most of technology (Webb, 2011). The importance cannot be over-emphasized of training teachers to effectively use ICT in the classroom and supporting curricula with effective ICT tools and learning platforms that enhance learning instead of replicating traditional instruction within a digital environment. Finally, since PISA does not use experimental designs, and usage data are based on students’ self-reports, thereby possibly introducing reporting biases, inferring cause-effect relationships and over-interpretation of the data should be avoided. In order to better understand the effects of usage on performance, further analysis of the types of activities undertaken by pupils as recommended in this article may shed light on how to improve ICT in education.

The international student assessment study known as Trends in International Mathematics and Science Study (TIMSS) moves beyond usage intensity to collect among other areas data on type of computer usage and student achievement for pupils in several developed and developing countries. For example in mathematics data are collected on computer usage for (i) exploring principles and concepts, (ii) looking up information, (iii) processing and analyzing data, and iv) practicing skills and procedures, while in science data are collected on computer usage for (i) looking up information, (ii) studying natural phenomenon using simulations, (iii) practicing skills and procedures, (iv) doing procedures and experiments, and (v) processing and analyzing data (Martin et al., 2012; Mullis et al., 2012). While country-by-country analyses of these usage patterns have begun, the relationships between use and outcomes are still unclear and require more systematic study.

At the national level, Brazil’s Regional Center for Studies on Information and Communication Technologies for the Development of the Information Society (Cetic.br), under the auspices of UNESCO, has been conducting specialized surveys on ICT aimed at the regular production of statistics on access to and use of information and communication technologies in different segments of society, providing important input for the process of formulating sector-based public policies. Started in 2010, one of these surveys is “ICT in Education,” which investigates the use of computers and the Internet in public and private schools (elementary and secondary) in urban areas of Brazil. The survey drew on work conducted by the International Association for the Evaluation of Educational Achievement (IEA) released in two publications: Sites 2006 Technical Report: Second Information Technology in Education Study and Sites 2006: User Guide for the International Database as well as the UNESCO Guide to Measuring Information and Communication Technologies (ICT) in Education from the UNESCO Institute for Statistics. Despite its limitations to an urban sample, the survey has become increasingly important for understanding the current scenario and trends related to the pedagogical use of new technologies and the Internet in Brazilian schools, especially in terms of the role of teachers as key agents for the dissemination of ICT in educational institutions.

This sampling of current collection instruments indicates the global shift that is occurring to better understand what learners and teachers are doing with technology, and forms a backdrop for recommendations to improve the UNESCO Institute of Statistics collection instrument.
Recommendations for the UNESCO institute of statistics data collection instrument

An analysis of literature as well as policies and practice from around the world were undertaken, which included analyzing the existing UIS indicators (UIS, 2014). From that basis the team developed research-based recommendations for strengthening the data collection post 2015. A panel of experts was convened at UNESCO headquarters in Paris in December 2015, which provided extensive feedback to the resulting new indicators draft during a four-hour critical review and discussion session, which further shaped and validated the framework. UNESCO selected 10 panel members as part their role in authoring additional white papers aimed at informing deliberations concerning global data collection by the UNESCO Institute of Statistics. The recommendations in this paper have been organized in sequence with the current Questionnaire on Statistics of Information and Communication Technologies (ICT) in Education (UIS, 2014). In each section, a brief research-based rationale is presented for the recommendations.

General information

The current survey probes five questions in relation to government policy, initiatives, incentives and equity. It is recommended that the updated survey:

- Collect new information on the response rate by total number of schools, by public and private sector, and compare the response rate in both categories in order to gauge the level of participation and to add validity to the results.
- Move the field labelled “main data source” into a relationship to each item to track its data source as needed.

Policy and curriculum

Ministers of Education are currently asked to provide data on policy and curriculum, including laws or regulatory mechanisms to promote the use of ICT in education and laws that address equity in favour of a number of disadvantaged groups, including females, minorities and rural people. The section also asks if there is a basic course on computer skills or computing, and which subjects (Mathematics, Natural Sciences, Social Sciences, etc.) have recommendations to use ICT to support teaching and learning. Finally, the section asks about the intended instructional time in basic computer skills and using ICT across the curriculum, and if there are accredited teacher education programs that include ICT-enabled distance education components. We recommend to add questions concerning:

- The development of leadership and training in ICT
- Emerging themes beyond basic computing involved in using computers to promote learning across the curriculum.

Recommended indicators for leadership and teacher training themes

Several authors and international organizations have emphasized having adequate infrastructure, technical support, and policies that encourage infrastructure usage by all students (International Society for Technology in Education, 2008; Kozma, 2003; Voogt & Knezek, 2011). For example, in their study of policies in Latin America and the Caribbean Hinostrosa and Labbe (2011) note that relatively few countries incorporated systems for evaluating policy implementation, half of the countries did not include enhanced student learning in policy, and twenty percent had not yet incorporated basic ICT competency into their curricula. Going beyond the basic conditions, educational systems must also take account of the changing context of knowledge, tools and practices of ICT that have impacted economics, science and culture across the world and have implications for education that extend from the science of learning through to the expectations of learners, and further to new horizons for curriculum and pedagogy (New Media Consortium, 2014). For example, the global shift toward the knowledge economy has created an acute need for deeper learning by larger and larger numbers of people, and so it is necessary to create and sustain ICT practices in education that support personalization, social community learning, acquisition of knowledge and expertise, and timely, effective formative performance feedback as outlined by the EDUsummIT forums (Gibson & Webb, 2013).

The EDUsummIT community involves 200 researchers, practitioners and policymakers from around the world who collaborated on the current status and research-based practices of ICT in education in nine theme areas since the program was established (Voogt & Knezek, 2013). The theme areas, which build on and extend the science of learning and global best practices in ICT in education into a systems view of education, have been studied and
validated by over 200 researchers from around the world since the early 2000’s (Voogt & Knezek, 2008) and are reviewed every two years. Results from this research and reporting are available at www.curtin.edu.au/edusummit. Defined below are recommended items with edited texts from the international EDUsummit reports:

School-community partnerships

New forms of partnerships are critical to new forms of schooling. The international dialog has evolved from a definition of public-private and informal-formal partnerships to a complex and evolving ecosystem of relationships of schools and society. Ministers and school leaders need to understand how this ecosystem responds to policy and funding as well as the progress of the science of learning and research on education. Davis, Eickelmann, and Zaka (2013) for example, indicate the relevance of considering the co-evolution of pedagogy and technology.

Mobile learning

There has been a gradual shift of understanding in the theory and practice of mobile learning in the last ten years, from a techno-centric perspective focusing on the attributes and affordances of the technology, to a learner-centred perspective focusing on the mobility of the learner (not just space and time, but also access to people and resources) and contexts (Kukulska-Hulme, Sharples, Milrad, Arnedillo-Sánchez, & Vavoula, 2009). One example of such a perspective is provided by Sharples, Taylor, and Vavoula (2007), who define mobile learning as “the process of coming to know through conversations across multiple contexts among people and personal interactive technologies” (p. 225). Mobile learning is also closely linked to informal learning, which is characterised by “personal ownership of codified knowledge, user-generated ideas, user-constructed contexts...personal and contextualised, and controlled by the learner” (Laurillard, 2009). Learner control and agency is thus at the heart of mobile learning and both personalised and collaborative learning opportunities can be afforded by mobile technologies. The field has thus begun to see development of theories of mobile learning (e.g., Sharples, Taylor, & Vavoula, 2007; Laurillard, 2007).

Educational equity

Digital divides exist between countries, including between girls and boys, women and men, rural and urban areas (McConnaughey & Sloan, 1995), young and old people (Becker, 2000; Fox & Madden, 2006), poor and rich people (Eamon, 2004), persons with or without disabilities, indigenous and “foreign” people, and “haves” and “have nots” (Resta, 2011). As stressed by van Dijk and Hacker (2003), the digital divide is a complex and dynamic phenomenon, one that is multifaceted. DiMaggio & Hargittai, (2001) suggested five dimensions along which the gender, age and socioeconomic inequalities may exist: (1) inequality in technical apparatus; (2) inequality in autonomy of use; (3) inequality in skill; (4) inequality in the availability of social support; and (5) variation in the purposes for which people use the technology.

Assessment

It has been recognised that assessment serves a range of formative and summative purposes. It is also clear that there are opportunities for IT-based assessments to serve 21st century learning goals including higher order thinking skills and deep knowledge (Gibson & Webb, 2013). The importance of assessment as a learning context has come to the fore and is particularly evident and arising in virtual performance assessment (Clarke-Midura, Code, Dede, Mayrath, & Zap, 2012; Pirnay-Dummer, Ifenthaler, & Spector, 2010; Webb & Gibson, 2015) where the experience of the assessment can be a learning engagement (Mislevy, Steinberg, & Almond, 2003). With the emergence of “big data” the need to develop assessment literacy (Stiggins, 1995) in teachers and other users has become even more important. Teachers need to understand the advantages and limitations of assessment types and processes and become confident in developing and analysing valid arguments from evidence (Black et al., 2010).
Creativity in the curriculum

This emerging research item refers to many forms of creativity, entrepreneurship, innovation, and non-verbal problem-solving, particularly when using ICT, that lead to new ideas and enhanced thinking skills (Mishra, 2012; Mishra, Cain, Sawaya, & Henriksen, 2013).

Indicators of ICT-enhanced teaching and learning

How teaching and learning is enhanced by ICT includes measures of several relevant domains; e.g. learning affordances for critical thinking, communications, creativity and collaboration (Kay & Greenhill, 2011), student characteristics, task and student performance characteristics, and evidence models (Mislevy, Steinberg, & Almond, 1999) to ascertain the extent, costs and benefits of ICT in teaching and learning.

Digital citizenship and cyberwellness

Parekh has written extensively on the topic of global citizenship (Parekh, 2008; 2003; 2002). Rather than arguing for absolute global citizenship, he suggests that “...citizens should be globally orientated, and able to discharge their duties to global others by exercising their responsibilities as democratic citizens and where necessary challenging nationalistic policies which are against the interests of mankind” (Parekh, 2003). This framework would allow the world’s citizens to move toward a global orientation; yet, within their region and nation-state contexts. Cyberwellness and keeping safe on the Internet is part of a larger conception of media and information literacy (UNESCO, 2013) which emphasizes the importance of accessing, evaluating and creating knowledge. Information literacy according to Catts & Lau (2008) includes capacities to (1) recognise information needs; (2) locate and evaluate the quality of information; (3) store and retrieve information; (4) make effective and ethical use of information; and (5) apply information to create and communicate knowledge.

Computer science and informatics in the curriculum

To what extent do schools give students learning opportunities in computer science and informatics? Computer science is the study of computation, coding, algorithms and related areas. Informatics is the use of information to solve problems (van Veen, Mulder, & Lemmen, 2004).

While these broad themes provide a foundation for new items that support the evolving paradigms essential for ICT in education, additionally learner-centred engagement and teacher use patterns are also integral to understanding the impact of ICT.

Learner-centred engagement in the curriculum

While students need access and support in order to learn with ICT, as important are the types of usage, for example to replace or extend traditional methods of teaching. In Latin America and the Caribbean, Hinostrosa and Labbe (2011) point out that the school and teacher’s perspective contributed significantly to the variation in learning opportunities from simple operations to creative uses of ICT. While it is therefore important to continue tracking the progress of countries on infrastructure, access and teacher training, a shift toward learner-centred usage is of critical importance for evaluating and comparing how students are being prepared for participating in the digital age. The shift toward learner-centred engagement has occurred at the policy level in developing countries for some years; however, it is evident that the implementation of such pedagogies has been fraught with inconsistency. For example, a case study in Namibia attempted to investigate the extent to which teachers (n = 145) were implementing learner centred approaches as outlined in reform policy documents. While teacher interviews suggested an understanding of the approach and most teachers reported to be implementing the policy in the classroom, the researcher reported that rote teaching was in fact the main method of instruction (O’Sullivan, 1999). Hinostrosa and Labbe (2011) make almost no comment on the types of pedagogical usage by teachers and the resulting learning opportunities for students in Latin America and the Caribbean, highlighting the need for new data collection measures and methods.
**Teacher use patterns**

Investigating what students are doing with ICT and which learning activities teachers plan for students when using ICT helps measure how ICT is being used in education. To better understand the impacts of ICT in education and increase the available information for comparison of various contexts, we recommend adding items concerning the usage patterns (e.g., time on tasks such as searching, working with models, creating, and computing).

Seminal work in seven countries by Stanford Research Institute (SRI) International (Shear, Gallagher, & Patel, 2011) demonstrated while ICT use in teaching is becoming more common, ICT use by students in their learning is still an exception in many schools. Findings from a study in Korea indicate that student teachers who held constructivist beliefs had strong computer efficacy, showed positive attitudes toward ICT in education, and were more interested in using ICT in future teaching practices (So, Choi, & Lim, 2005). A more recent ICILS study also confirms that teachers with greatest use are those who are more confident in their abilities, and those who develop higher levels of confidence in using ICT in teaching tend to work in school environments where staff learn together, collaborate in institutional planning, and where there are fewer resource limitations concerning the use of ICT in teaching and learning (Fraillon et al., 2013).

**Government expenditure**

The current ICT in education survey instrument attempts to collect data on the total expenditure assigned to ICT infrastructure, including both hardware and software. We recommend that this section be removed from the current survey and be included in the main UIS survey on expenditures. Having a few financial items divorced from the main expenditure survey is not the best option since responding to a few items in the ICT in education survey is hard to respond to resulting in poor quality data. However, we recommend for that survey to include expenditure on teacher professional development.

**ICT Infrastructure**

The current survey asks countries to indicate which education institutions have access to a broad range of infrastructure items, from basic electricity to broadband internet access. The ICILS study found that students from countries with greater access to computers in schools tend to have stronger computer and information literacy skills (Fraillon et al., 2013). The ICT infrastructure paper by Twining and Davis (2015) commissioned by UNESCO-UIS also addresses infrastructure issues in depth, so the following bulleted recommendations are offered as additional guidelines for survey-based comparisons.

**Deployment patterns by space and type of access**

In order to understand the usage of ICT, it is recommended that data be collected in terms of “minutes per week” estimated by the teacher (and even students in the classroom where possible) that focuses on deployment patterns inside and outside the classroom to help triangulate and validate deployment patterns. Documenting the access mode type (e.g., mobile, wireless or wired connections) would be beneficial. By collecting information on where and via what modes students are accessing ICT, comparability among schools will be improved and impacts can be better understood.

**Unique student data records**

To support analyses and understanding of the impacts of ICT on students, schools and systems, we recommend the creation of a unique student identifier, ideally that is unique within the largest system of data collection envisioned (e.g., unique in the country). The unique student identifier (an integer or integer plus alphanumeric student ID code) is recommended to allow in-depth analytics at the most granular level possible (i.e., individual student) and to prevent duplication of data. If these fields exist, then fine-grained analyses of gender equality issues are supported, otherwise gender and other issues are difficult to detect and validate.
Data elements for tracking specific impacts of ICT policies

ICT infrastructure includes data system structures for documenting how key policies drive the impact of using ICT in teaching and learning. Thus, a series of policy level estimations is recommended concerning the movement of schools on three measures that relate to the impact of ICT on teaching and learning: Retention, Student Satisfaction and Learning Analytics. The survey should ask what percentages of schools are in the following categories: (1) No policy, (2) Policy developed, (3) Implementation planned or piloted, (4) System fully implemented. The fields would be defined as mutually exclusive. “Policy developed” means that its implementation has not yet been planned or piloted. “Implementation planned or piloted” moves a school out of the “Policy developed” category, “System fully implemented” moves a school out of the category “Implementation planned or piloted.” Each line adds to 100% of the country’s schools.

ICT Tools

ICT tool allocation by pedagogical use

The types of ICT used in a school are closely aligned with the expected outcomes in that context (Dede, 2008; Vanderlinde, Braak, & Dexter, 2012). Key uses of ICT in schools can be divided into four types: ICT used for learning, ICT used for content or skills assessment, and ICT available for use during non-class time. One way to measure the amount of use in these categories is to count the number of available devices or the student-to-device ratio for these purposes. It is therefore recommended that ICT device allocation numbers be subdivided to provide data on learning, content assessments, ICT skills assessments and devices allocated for free time use. The numbers would not be mutually exclusive, but give an approximate student-to-computer ratio for each of the four types of use. An alternative measure might be to ask the school administrator to state the student-to-computer ratio available for each activity type.

Emerging paradigms of tool use in ICT in education

The Horizon Reports of the New Media Consortium NMC (2013; 2014) provide a source for comparing ICT uses being contemplated by ICT leaders in education. For over a decade these yearly reports detail the six technologies that will soon impact colleges and universities. The method of the report is based on a survey of current ICT leaders, who provide information about what is on their immediate timeline for adoption and implementation. Thus the reports give a field-based prediction about worldwide directions being taken in ICT in education. It is recommended that (1) Emerging paradigms in the use of ICT tool-based approaches to delivery and learning be tracked for comparison and change over time; (2) Estimate the number of years until a policy will be in place and until implementation will begin. Emerging technologies in recent NMC reports include bring your own device, cloud computing, flipped classroom, games-based learning, learning analytics and the internet of things.

Enrolment

Gender and cultural access to ICT

Enrolment statistics play an important role in promoting equal access to knowledge by all sectors of a society including equal opportunities for females and children from all different cultures, regions (urban versus rural), and socio-economic levels. Gender-sensitive indicators and analysis are valued in national policies and to provide evidence-based information for concrete gender-specific measures such as projects and programmes (UNESCO, 2013). Items are recommended that (1) Use robust methods to identify gender, cultural and other biases; (2) Include gender and socioeconomic or cultural analysis of participation in online learning.

This latter recommendation recognizes that massively open online learning (MOOCs) can be valid forms of learning and need to be acknowledged (DeFreitas, Gibson, & Morgan, 2015). MOOCs are scalable, free sources of information that are open and available to anyone with access to the Internet. National policies are needed that review and make use of these unique Open Educational Resources to raise the level of education for all people. Schon and Conole (2014) provides a special issue on quality in MOOCs.
Teaching staff

ICT competency framework for teachers

A recent UNESCO paper by DuToit “Teacher training and usage of ICT in education: New directions for the UNESCO Institute for Statistics global data collection in the post 2015 context” presents the UNESCO ICT Competency Framework as a guide to professional development of teachers, which suggests that six dimensions of training be provided at three levels. The six dimensions that we recommend new items begin to track include policy, curriculum and assessment, pedagogy, ICT as tools, organization and administration, and professional learning and the levels are defined as technology literacy (awareness and basic knowledge), knowledge deepening (applying knowledge) and knowledge creation, including self-management and sharing knowledge as a model learner.

Conclusion

In order to understand the use of ICT in education in today’s society, the indicators of evolving learning paradigms need to focus on what teachers and students are doing with ICT in relation to teaching and learning. Usage patterns will be affected by and can be correlated with ICT infrastructure, access to technology and information within the curriculum, and teacher training and support for the implementation of ICT.

The proposed international survey changes recommended in this article distribute the responsibility for reporting between the ministerial representative, the school administration, and teachers with evidence from students. Policy is led by and reported by the minister, access to tools is reported by the school administrator, and the balance of items are reported by the teacher, with a subset of classroom learning opportunity items confirmed by students.

A limitation of this current research is that the recommendations have not been field-tested. Next steps for future work include refinement of the items, field-testing of the new data collection methods, and if appropriate, adoption into the UIS data collection processes and instruments. The authors stand ready to join in the conversation and contribute to the development of ideas, and we hope that we have provided food for thought for others involved in shaping international data collection to better meet the evolving learning paradigms afforded by ICT in education.

The recommendations are proposed for the purpose of addressing the evolving mission, methods and core principles of ICT in education; the role if ICT in hastening the emergence of learner-centered pedagogies; and to document the types of learning activities and usage patterns that best support full participation in the knowledge society.

References


Are Games Effective Learning Tools? A Review of Educational Games

Sara de Freitas
Birkbeck College, University of London, Malet Street, London, United Kingdom // sara@cs.bbk.ac.uk
(Submitted December 7, 2016; Accepted January 9, 2017)

ABSTRACT
The literature around the use, efficacy and design of educational games and game-based learning approaches has been building up gradually and in phases, across different disciplines and in an ad hoc way. This has been problematic in a number of ways and resulted in fragmented literature and inconsistent referencing patterns between different sub-disciplines and countries. This is mainly because no distinct single-disciplinary perspective has emerged because of: the cross-disciplinary nature of educational games, a reliance on single-disciplinary contexts for studies, changing terminologies in different contexts and the use of multi-methodological approaches. Distinct perspectives from education science, game science, neuroscience and information science have deepened our understanding of play and games. This research has become more quantitative, rigorous and nuanced as a result of more studies focused upon therapeutic health applications of games, the serious games research movement and more efficacy and comparative studies that examine and quantify utility.

Keywords
Educational games, Serious games, Game science, Neuroscience and games

Introduction
Defining efficacy in educational contexts can be challenging due to the range of variables involved in different learning contexts. Additionally, there are disciplinary restraints that have traditionally meant that cross-disciplinary approaches to data collection and analysis have been broadly discouraged. However to understand education, and in particular questions around efficacy, necessarily we need to adopt more cross-disciplinary approaches. As an example, research emerging from education science is being supplemented by findings from computer science (e.g., interfaces and interactivity), neuroscience (e.g., brain function and activity) and information science (e.g., analytics and user-modelling). Notably these include findings from computer science which allow us to consider usability improvements and human-computer interactions (e.g., Barr et al., 2007), findings from neuroscience which provide a greater understanding of how games impact our brain plasticity (e.g., Bavelier et al., 2012; Kühn et al., 2011; Kühn et al., 2014) and approaches that use analytics in games as a replacement for assessment (e.g., Serrano-Laguna et al., 2012). Together, these findings help provide a broader understanding of how we can model learning experiences in digital, data-rich game environments, and tell us more about how we learn.

The review found that “game science” is emerging as a new term to replace “serious games” which has been a significant term for the game studies research community for the last ten years. Similar to “serious games”, the new term aims to link game studies to a greater scientific capability which has the potential to help us model and better understand: the learning behaviours of individuals and groups in game environments, learning design through the metaphor of game design and how games and play work to help people learn.

Establishing the efficacy of games and learning is a complicated endeavour. It needs to be kept within a wider context of understanding how we learn. So how game science fits into the wider disciplinary framework is a critical consideration. When viewed from this educational perspective, the notion of “game science” is part of the field we might call, “education science” and due to its digital nature it is often placed within the sub-discipline of Technology-Enhanced Learning (TEL). However, clearly there is important work to be found across a range of different areas including: human-computer interaction (e.g., Barr et al., 2007), health education and research (e.g., Papastergiou, 2009), neuroscience research (e.g., Kühn et al., 2011; Colzato et al., 2013; Lewis, 2013; Kühn et al., 2014), and across other literature such as business and management (e.g., Pasin & Giroux, 2011), school education (e.g., Hainey et al., 2016), advertising and marketing (e.g., T精灵 & Capella, 2013), military training and simulations (e.g., Hassain et al., 2012), environmental awareness-raising (Rebolledo-Mendez et al., 2009), therapy training (Horne-Moyer et al., 2014), teacher training (e.g., Kenny & McDaniel, 2011) and emergency-response training (e.g., Chen et al., 2008). One challenge with the literature so scattered is that not all researchers acknowledge the breadth of the area and range of applications, and therefore miss vital academic contributions by looking too narrowly at the literature-base. The situation is exacerbated by rifts between US and European research in serious games and between simulation and games literatures, and often you can see researchers will completely ignore critical papers from one “side” of the Atlantic or the other leading
to misunderstandings and incomplete starting points. Despite a number of special issues on games crossing several fields, the continuation of fragmentation of the field has again happened with the split between researchers in serious games and the new area of “gamification”. Gamification here is used to mean the application of elements of game-mechanics and/or game-design techniques. To attempt to bring the literatures closer together and to attempt to harmonize some of the terminology, this paper aims to map out the potential new ground for learning as evidenced in the sub-field of technology enhanced learning that defines game-based learning approaches.

To overcome these significant disciplinary challenges, this paper seeks to outline some of the major contributions of the field from different disciplines over time and synthesise these using an integrative approach to a broader education science perspective. The aim is to problematize the current scope of Education Studies and to reposition game science more critically within this educational context and perspective.

Methodology of literature review

This article has adopted a “grounded theory” approach used over a number of years to assess the main themes emerging from the fields that touch on educational games. The method used included a semi-systematic review process with a single-coder, wide literature searches across databases using keywords to collect high impact and cited articles and is supplemented with a journal hand-search. Keywords included educational games, serious games, learning games, web-based games and digital games. Once key texts were identified from the literature search, these were grouped into disciplinary perspectives. The emerging perspectives of education science, game science/studies, neuroscience and information science were distilled and key articles identified were included in this review.

The Review: A recent history of game science

Wave 1: What are games?

Some of the earliest work in the field of game science focused upon, changing definitions and nominations of educational games. For example, work that outlined classifications of games, typologies and ontologies was found in the early literature (e.g., Caillois & Barash, 1961; Sutton-Smith & Roberts, 1971). While the earlier work focused upon structuralist perspectives upon educational games as consistent with the trend for semiotics and structuralist analysis, the theme re-emerged later on in the more recent literature as a theme of consideration (e.g., Elverdam & Aarseth, 2007; Kamii & DeVries, 1980; Salen & Zimmerman, 2004). But the more consistent theme of poststructuralism and postmodernist perspectives necessarily focused more upon notions of play than structure also in line with constructivist and qualitative studies.

It is perhaps ironic that constructivist approaches to learning have become so associated with qualitative approaches as although the work does focus upon individual construction of meaning, the social constructivism of Vygotsky (1980) and others does propose learning in social groups as a central component of learning. But here a split between the American and other literatures can be noted as a de-emphasis of social learning and a greater focus upon Skinnerism and behaviourist approaches as consistent with the individualism and competition of the American ideal. The mode of bringing education theory together with an American individualist twist and its bringing into the paradigm of psychology jointly ensured that the more social focus emphasised by Russian theorist Vygotsky did not become the dominant discourse. The legacy of this can also be seen in the more general sparseness of social learning theory and was compounded by difficulties with researching and analysing group work, a trend that is partially being reversed in studies such as Star where collaboration rather than competition techniques are emerging (Star, 2015).

Wave 2: The serious games movement

Negative publicity around violence in games, in particular entertainment games have attracted popular attention. The robust evidence of games causing violence has overall been inconclusive (e.g., Elson & Ferguson, 2015) – but nonetheless the distinction between games for entertainment and games for non-entertainment was a major driver for why the “serious games movement” occurred in the early 2000s (Blumberg et al., 2013). However, once non-entertainment games could be demonstrably “taken seriously” for purposes such as military training and health education and therapy then the research field gained greater credibility.
Early “serious games” titles, such as America’s Army, have set the bar high in terms of the budget ($33 million invested up until 2015 in all titles). Although small budgets next to entertainment games, (e.g., $265 million for Grand Theft Auto 5), America’s Army is still considered one of the best exemplars of a serious game today. Having been first published in 2002, it has 13 million registered players who have played 260 million hours. Developed by the US Moves Institute to solve the recruitment problems of the US Army, unfortunately the game has proved to be more of an oddity than a trend. Few large budget serious games have been developed since 2002, and those that have been commissioned have not always enjoyed longevity of support once piloting phases have concluded, e.g., Code of Everand (Dunwell et al., 2014). During this period, although relatively disconnected from the mainstream games literature, the “serious games movement” did gain important contributions from game studies, such as a deeper understanding of the mechanisms of competition as a design component (Cagiltay et al., 2015), how to balance entertainment principles of fun with instructional design and the need to integrate teams of developers, writers and instructional designers.

Wave 3: Technology-enhanced learning perspectives: Out of the wilderness?

The next phase of focus upon educational games borrowed heavily from technology-enhanced learning approaches. There, a focus upon verification and validation of online learning and e-learning was leading to a wide range of comparative learning studies. Again studies were often lacking in robust methodologies, but were beginning to seek a more scientific basis for analysing the efficacy of learning techniques. This approach was driven-out of concerns about the quality of learning in online settings and studies were often more utility-focused. While the early studies had attempted to group games in typologies and genres, these studies focused upon comparisons with other e-learning formats and against traditional learning measures (e.g., Knight et al., 2010).

Out of this work, a movement to understand game design emerged, how could games be designed for different learning contexts? Could commercial off-the-shelf (COTS) games be used? These questions led to a range of studies of games in educational contexts and collections of case studies (e.g., Kim et al., 2009; Michael & Chen, 2005; Prensky, 2005; Shute et al., 2009). This phase of research was dominated by educational perspectives.

However, there were significant difficulties with uptake of games in educational contexts. As Simon Egenfeldt-Nielsen outlined in his thesis (Egenfeldt-Nielsen, 2005), games did not fit into the one-hour lessons, into the single disciplinary focus or into the single-teacher model of traditional learning. Games were disruptive, they demanded greater changes to the traditional delivery and infrastructure of education in schools, colleges and universities. Beyond traditional learning paradigms (see Table 1), game-based approaches required: cross-disciplinarity, longer class durations, mixed student groups, social learning and team-teaching models to come into place to really capitalise on the merits of the game and gameplay as learning approaches (de Freitas, 2014).

Four disciplinary perspectives from the literature

While it is difficult to be too prescriptive with the time periods, the research does seem to fall broadly into four broad disciplinary categories: education science including theory and practice studies and using elements of pedagogy and psychology, game science contextualised through technology enhanced learning, neuroscience that have focused upon brain-function and plasticity and information science-driven studies that focus more upon data analytics and behavioural modelling. The following sections outline these perspectives (see summary in Table 1):

Education science perspective on educational games

Major contributions to understanding learning formed early theoretical and developmental approaches to learning. Through understanding learning as cognitive and developmental sets of processes, theorists and educationalists, such as Jean Piaget, defined ages and stages of development associated with “normal patterns of development” (Piaget, 1971). But Piaget also understood the importance of play in learning (Piaget, 2013). Play has been a theme of the work around games necessarily, but has not been a well-understood aspect of learning. More recent play research by Jean Twenge and others shows how important and developmental play is to learning (e.g., Campbell & Twenge, 2015; Chudacoff, 2007; de Freitas, 2014; Gray, 2011; Twenge & Campbell, 2008).
In the light of the internet, broadened connectivity and mobile access to online educational content, there has been a de-emphasis on content and curriculum and a sharpened focus upon digital literacy and 21st century skills. Employability for the changing global employment market presents new needs for graduates and students (Harlow & Bowman, 2016). The move to a more utilitarian position, driven by education via web-based technologies and digitisation, has reworked how we deliver a university education and even challenged what the role of the university is (Sugden, 2013).

<table>
<thead>
<tr>
<th>Traditional paradigm of learning</th>
<th>New learning paradigm</th>
<th>Future learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curriculum-based pedagogy</td>
<td>Challenge and activity-led learning</td>
<td>Student developed pedagogy</td>
</tr>
<tr>
<td>Tutor-led learning delivery</td>
<td>Peer-focused interactions</td>
<td>Artificial Intelligence (AI)</td>
</tr>
<tr>
<td>Classroom and lecture hall focus</td>
<td>Any-time, anywhere learning</td>
<td>Seamless lifelong learning</td>
</tr>
<tr>
<td>Summative assessments</td>
<td>Formative assessment / Peer assessment</td>
<td>No assessments / levelling, points and awards</td>
</tr>
<tr>
<td>Age and stage</td>
<td>Competency and personalised learning</td>
<td>Unique learning patterns</td>
</tr>
<tr>
<td>Text-focused</td>
<td>Multimedia usage</td>
<td>Adaptive learning</td>
</tr>
<tr>
<td>Traditional curriculum e.g., literacy and numeracy</td>
<td>New curriculum e.g., 21st century skills</td>
<td>Hidden curriculum e.g. personalised skills and cognition training</td>
</tr>
<tr>
<td>Core curriculum</td>
<td>Work readiness</td>
<td>Blended work and learning</td>
</tr>
</tbody>
</table>

In the author’s recent work, she articulates this disruption as a “new learning” paradigm. One that focuses upon problem-, challenge- and active pedagogy, peer learning and is competency-based and personalised (de Freitas, 2014). This differs from the traditional modes of curriculum-based and tutor-led approaches. With the work on games we can begin to see the rudiments of what the author calls a “future learning” paradigm, which advances to student-led approaches where adaptive learning is scaffolded through AI bots, assessment gives way to in-built levelling-up and the curriculum is hidden (See Table 1).

### Game science perspective upon educational games

One of the main stated inhibitors to uptake of educational games and approaches was the lack of robust scientific and evidence-based research. The first randomised and pragmatic randomised controlled trials (RCTs/PCTs) started in the late 2000s. One of the early trials was undertaken by Knight et al. (2010), focusing upon a comparison between traditional and game-based approaches in emergency response training. Arnab et al. undertook an RCT which considered a serious game in a classroom setting. Miller and Robertson undertook an RCT on educational benefits of games consoles in classrooms (Miller & Robertson, 2011). While Star considered a randomised control trial for gamification in StarQuest to identify cooperative and competitive design elements in university students (Star, 2015). Arbogast et al. (2014) were examining the use of an educational game for road crossing in their recent study.

Unsurprisingly most recently RCTs involving games have focused upon health and medical conditions including patients with weight conditions (e.g., Ahola et al., 2013; Maddison et al., 2011; Siervo et al., 2013; Straker et al., 2011; Straker et al., 2013). Fung et al. (2012) considered the use of the Wii Fit for knee rehabilitation. Foss et al. (2013) used their randomised control trial to discover effective use of the i-Bit which is a novel binocular device which uses games and videos to improve patients with a lazy eye. Picherri et al. (2012) looked at the impact of a dance game upon gait. Another popular area for study was the impact of games upon the elderly. An interesting study by Nouchi et al. (2012) explored the positive impact on executive and processing speeds on the elderly of brain training games in their study. While Mayas et al. (2014) explored the plasticity of the brain in the elderly after non-violent game play. A study on Wii Fit games for patient’s living with Parkinson’s disease was undertaken recently by Pompeu et al. (2012); and one looking at improvements from gameplay with Diabetes sufferers (Kempf & Martin, 2013). Allam et al. (2015) in their RCT on gamification in an online intervention for Rheumatoid Arthritis Patients found that “physical activity increased over time for patients having access to social support sections plus gaming (unstandardised beta coefficient $\beta = 3.39, p = .02$).” Patients were also more empowered and used services less as a result.

In addition to more quantitative studies such as RCTs/PCTs, meta reviews have offered important research contributions to overcoming the prevalence of different disciplinary perspectives. Often these reviews have been
cross-disciplinary in scope and dimension, single topic-focused, centred-upon comparative studies or in support of game design improvements. While there was a large group of studies done on violence in games (e.g., Anderson & Bushman, 2001; Anderson et al., 2010), these studies do not have much relationship with educational games which do not use violent metaphors or gameplay. One of the earliest meta-reviews was undertaken by Randel et al. (1992) considered literature 1963-1984, finding that of the 67 studies undertaken over the period, “38 show no difference between games and conventional instruction; 22 favour games; 5 favour games, but their controls are questionable; and 3 favour conventional instruction.”

Vogel et al. (2006) in their review included simulations and games, it found that “games and interactive simulations are more dominant for cognitive gain outcomes,” it also found that when students were empowered to control access to simulations and games there were significant advantages over when access was tutor-controlled, when no advantage was found. Ke (2009) undertook his meta-review in 2009, reviewing 89 studies finding that there was a need for more longitudinal studies, less segmentation in the literature and more empirically-based studies. 65 out of 89 studies evaluated the effects of the game upon learning. From the empirically-based studies 34 out of the 69 found positive outcomes from using games, 17 had mixed results, 12 reported “no significant difference” with traditional instruction approaches – and one study found traditional methods more effective.

Wouters and Van Oostendorp (2013) undertook a meta-review on instructor-support in game environments, finding that “instructional support in game-based learning environments improved learning,” further that greater improvement was found in skills-based learning. Wouters et al. (2013) also found in another meta-analytic review of literature that “serious games were found to be more effective in terms of learning (d = 0.29, p < .01) and retention (d = 0.36, p < .01), but they were not more motivating (d = 0.26, p > .05) than conventional instruction methods.” This refuted much of the educational literature that had found games to have strong motivational gains (e.g., Garris et al., 2002). The study also found that “learners in serious games learned more, relative to those taught with conventional instruction methods, when the game was supplemented with other instruction methods, when multiple training sessions were involved, and when players worked in groups.” Sitzman (2011) found that “consistent with theory, post-training self-efficacy was 20% higher, declarative knowledge was 11% higher, procedural knowledge was 14% higher, and retention was 9% higher for trainees taught with simulation games, relative to a comparison group.” However she did find evidence of publication bias.

Connolly et al. (2012) undertook their meta-review, in contrast to Wouter et al. (2013), they found improvements in motivation. Their study reviewed 7,392 papers in total and found that “playing computer games is linked to a range of perceptual, cognitive, behavioural, affective and motivational impacts and outcomes.” Of the most recent reviews undertaken since 2014, Clark, Tanner-Smith and Killingsworth (2015) have found “results from media comparisons indicated that digital games significantly enhanced student learning relative to nongame conditions (Formula = 0.33, 95% confidence interval [0.19, 0.48], k = 57, n = 209).”

Neuroscience perspective on educational games and play

Our understanding about how we learn, through brain science and experiment, largely builds upon work of Edelman (1987) and Kandel and colleagues (2000). The specific scientific studies of neuroscientists Daphne Bavelier and Simone Kuhn have helped to shape the field and given great insights into the power of games to support advanced learning. Greater brain volume and plasticity with gameplay (Kühn et al., 2011; Kühn et al., 2014) and greater transferability of skills such as hand eye coordination, memory abilities and visual acuity (Green & Bavelier, 2003; Green & Bavelier, 2008; McDermott, Bavelier & Green, 2014) are amongst the more important findings revealed in recent studies. For example, Green and Bavelier (2008) undertook a review on brain plasticity and learning. They concluded, “possible characteristics of training regimens are proposed that may be responsible for augmented learning, including the manner in which task difficulty is progressed, the motivational state of the learner, and the type of feedback the training provides. When maximally implemented in rehabilitative paradigms, these characteristics may greatly increase the efficacy of training” (Green & Bavelier, 2008, p. 699).

Beyond these studies, it is hoped that we will begin to answer some questions, such as why are games effective learning tools? How can games be used to model social learning behaviours?
Information science perspective on educational games

One of the recent game-changers in the field of education research has been access to large datasets gleaned from learning management systems (LMS), student information systems (SIS), interactive environments and other computer-generated environments, such as digital games. In digital environments, such as games, all data can be collected and analysed relatively easily (Deterding et al., 2015; Loh et al., 2015). In these more data-rich environments the possibility to look at social learning behaviours has emerged (e.g., Gentile et al., 2009; Steiner et al., 2015). The study of social learning behaviour allows for individual and cohort mapping, comparative cohort studies and importantly longitudinal studies. The richness of learning data – or learning analytics – has led to more quantitative and longitudinal studies that involve large student populations (e.g., de Freitas et al., 2015) to supplement the preponderance of qualitative studies. This recent focus upon quantitative study of learning has real potential to inform how we design “new learning” and ensure that our students are suitably engaged and actively partnering in their learning. This is a powerful capability, but not without complex ethical issues in terms of privacy, de-identification of data, informed consent, data management and archiving (e.g., Pardo & Siemens, 2014; Slade & Prinsloo, 2013), some of which may be overcome in time by blockchains (Sharples & Domingue, 2016). Work is needed to ensure that feedback systems are beneficial to the attainment and success of the learner whilst enshrining ethical considerations and transparent approaches.

The notion of game analytics brings together large datasets for analysing human behaviour, supporting learning experiences and supporting individual and group performance and personalisation capabilities (e.g., El-Nasr, Drachen & Canossa, 2013; Drachen et al., 2013; Gibson & de Freitas, 2016).

Table 2. Contributions to game science from four disciplinary perspectives

<table>
<thead>
<tr>
<th>Contribution from education science</th>
<th>Contribution from game studies/science</th>
<th>Contribution from neuroscience</th>
<th>Contribution from information science</th>
</tr>
</thead>
<tbody>
<tr>
<td>Importance of play to learning has been confirmed in play studies e.g., identification of importance of play (Piaget, 2013)</td>
<td>Game Studies and Science literature includes insights such as increased motivation (e.g., Star, 2015; Plass et al., 2015; Attali &amp; Arieli-Attali, 2015)</td>
<td>Greater brain volume and plasticity with game play (Kuhn et al., 2011; 2014)</td>
<td>Data modelling will allow us to map human behaviour more closely by using data interactions in games (e.g., Gibson &amp; de Freitas, 2016)</td>
</tr>
<tr>
<td>Longitudinal studies of examining play patterns (e.g., Twenge &amp; Campbell, 2008)</td>
<td>Pragmatic and randomised trials have confirmed that games can be more effective learning tools than traditional modes (advance on e-learning which found no significant difference with traditional modes) (e.g., Knight et al. 2010; Miller &amp; Robertson, 2011; Straker et al., 2011)</td>
<td>Greater transferability of skills such as hand eye coordination and visual acuity (Bavelier, 2003 (with Green) and 2014 papers)</td>
<td>Analytics allows for personalisation in games (e.g., El-Nasr, Drachen &amp; Canossa, 2013; Drachen et al., 2013)</td>
</tr>
<tr>
<td>How patterns of play can impact learning (e.g., Chudacoff, 2007; Gray, 2011)</td>
<td>Use of combined measures introduced including qualitative and quantitative measures (e.g., Kato et al., 2008)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Discussion and conclusions

This review has aimed to reposition the emergent game science area of research within four inter-related disciplinary contexts of: education science, game studies, neuroscience and information science literatures.

Key challenges for integrating the research base are summarised as:
• The literature is so scattered across different disciplines that not all researchers acknowledge the breadth of the area and range of applications, and therefore miss vital academic contributions by looking too narrowly at the literature-base.
• Beyond traditional learning paradigms (see Figure 2), game-based approaches require: cross-disciplinarity, longer class durations, mixed student groups, social learning and team-teaching models to come into place to really capitalise on the merits of the game and gameplay as learning approaches, which are problematic for traditional formal education systems to incorporate.
• Work is needed to ensure that feedback systems used in educational games are beneficial to the attainment and success of the learner whilst enshrining ethical considerations and transparent approaches.
• Finding the balance between game playability and fun and solid learning design that aligns learning outcomes with assessments (in-game or as part of the blended experience) is a key challenge for effective educational game design.

The overall findings of the studies confirm that a more robust literature-base has grown considerably in recent years and has led to the notion of “game science.” Moreover, while the efficacy of educational games is hard to measure, findings from quantitative RCT and more data-driven longitudinal studies are giving us more robust findings to build and improve design of learning experiences, involving gamification and game-based elements and enhancing student success. What we have learnt from the research as well is the importance of using combined measures including qualitative and quantitative measures (e.g., Kato et al., 2008).

Game science is emerging as a robust and dynamic area of research crossing several disciplinary areas and redrawing the scope and research questions that intersect with learning efficacy and design. The future of this sub-field might include bringing together the substantive literatures of simulations, serious games, gamification and education technology. The two issues of cross-disciplinarity and methodology will be key for establishing the lines of the discipline, with the absorption of randomised controls, meta reviews and large dataset analyses combining with the qualitative methods established in education such as content analyses, case studies and ethnologies and with other approaches such as neurological studies and social network analyses to provide a level of granularity that supports better learning design and an improved student experience, through modelling social behaviours.

To the question: are games effective learning tools, the answer from the research is overwhelmingly positive. Going further, the weight of the research findings seems to point to significant improvements in game over traditional methods, and these are further enhanced by blended approaches that utilise game and face-to-face approaches. The work distilled from RCTs is particularly positive and indicates that educators are now challenged with the best ways to implement game-based approaches in their institutions. While it seems that games do enhance student motivation, are engaging and can be associated with behavioural change, more active design studies are needed to ensure that the best interests of the learner are met in different contexts. As educational games enter into a new wave of implementation, it will be interesting to see whether the lessons from across the different disciplines are absorbed into general practice.

It is clearly a challenge for educational institutions, policymakers and practitioners, but with the growing evidence-base advances in quality and overcoming challenges of privacy and design might be forecast. Despite resistance to the adoption of game-based approaches in schools, colleges and universities, like online learning, it will be a matter of time before the cost benefits drive uptake widely and the full implication of the research is fully understood. As the traditional learning paradigm gives way to the new learning and then on to the future learning approaches, game-based learning will become more embedded into practices, be personalised and hide the curriculum in more seamless ways. But researchers, policymakers, managers and practitioners in the field will need to work hard to ensure: distillation of key benefits, join-up of the literatures, harmonising different disciplinary perspectives, methodological challenges and creation of a shared terminology between these four disciplinary perspectives.

References


80


Teaching Classical Mechanics Concepts using Visuo-haptic Simulators

Luis Neri*, Julieta Noguez, Victor Robledo-Rella, David Escobar-Castillejos and Andres Gonzalez-Nucamendi

Engineering and Sciences School, Tecnologico de Monterrey, Mexico City, Mexico // neri@itesm.mx // jnoguez@itesm.mx // vrobledo@itesm.mx // A01170737@itesm.mx // anucamen@itesm.mx

*Corresponding author

(Submitted October 12, 2016; Revised February 11, 2017; Accepted March 17, 2017)

ABSTRACT

In this work, the design and implementation of several physics scenarios using haptic devices are presented and discussed. Four visuo-haptic applications were developed for an undergraduate engineering physics course. Experiments with experimental and control groups were designed and implemented. Activities and exercises related to classical mechanics concepts were applied in virtual scenarios. The importance of carefully designing haptic scenarios and planning the implementation process to foster greater student engagement was emphasized. The quality of the visualization and friendlier interaction with bodies in the simulation are essential factors. The haptic intervention was evaluated through a perception questionnaire about the use of the visuo-haptic simulators. The results strongly indicate that most students were motivated to use haptic technology. In addition, post-tests were administered to compare the potential learning gains of the experimental and control groups. The results indicate that students in the experimental group obtained greater learning gains than those in the control group in two scenarios. These findings suggest that, by incorporating properly designed visuo-haptic learning environments, students are engaged and motivated to learn physics concepts, thereby supporting better learning outcomes.

Keywords

Haptic devices, Virtual simulators, Technology-enhanced learning, Physics concepts, Active learning

Introduction

Traditionally, physics concepts have been conveyed to different audiences, including students, through lectures. In educational institutions, this approach is considered an effective learning method (Vanlehn, 2006). However, as science and technology constantly evolve, new learner-centered solutions have been developed. Several types of simulations, such as virtual laboratories and learning environments, have been incorporated in teaching. Such simulations provide alternatives for students to improve their skills and learning processes (Yang & Zhang, 2011). Consequently, such virtual environments have established the area known as technology enhanced learning or e-learning.

Studies have identified that the use of haptic technology is a useful approach to enhance interaction (Laycock & Day, 2003). By interactions between the user and virtual objects, the force feedback of haptic devices can recreate the sense of touch. Therefore, such haptic devices have enabled the development and creation of new interactive applications (Figure 1). Haptic technologies have been used in various areas, such as navigation, education, e-commerce, gaming, and arts (Krenek, 2001; Ni, Krzeminski, & Tuer, 2006; Benes et al., 2006; El Saddik, 2007; Abdul-Massih et al., 2011; Liang et al., 2011; Minogue & Jones, 2006). In addition, visuo-haptic simulators have been implemented for teaching and training in the medical field (Coles, Meglan & John, 2011; Basdogan et al., 2004; Ricardez et al., 2014; Escobar-Castillejos et al., 2016).

In traditional physics courses, students often find it difficult to visualize and understand physical phenomena. This difficulty increases when abstract and non-tangible concepts are involved, such as electromagnetic interactions (Sanchez et al., 2013; Neri et al., 2015; Reiner, 1999; Dede et al., 1999). As alternative solutions, real laboratories and virtual simulators have been used to help students understand the nature of physical interactions (e.g., Zacharia and Olympiou, 2011). However, not all educational institutions have laboratories or the equipment required to perform suitable experiments. Virtual simulators can be a substitute for real laboratories (e.g., the PhET Interactive Simulations; PhET, 2018); however, most virtual simulators focus on visualizations, although some include the sense of hearing (Han & Black, 2011). Thus, haptic devices have the potential to improve physics education by adding tactile sensations.

In this paper we explore how haptic devices can be used to effectively enhance the learning experience for traditionally difficult classical mechanics concepts. An interactive simulator with four scenarios was developed. The experiments were displayed on a screen and controlled using a haptic device. The main objective of this study was to assess whether the use of haptic devices combined with appropriate virtual environments can
engage and help students understand the nature of some mechanics concepts. Consequently, a user study was performed to obtain student perceptions about the use of these simulators. In addition, the knowledge students acquired working with the visuo-haptic simulators was evaluated.

![User interacting with a visuo-haptic environment](image)

Figure 1. User interacting with a visuo-haptic environment

The remainder of this paper is organized as follows. First, visuo-haptic applications that focus on teaching physics concepts are reviewed. In the experimental design section, the application architecture, the main features of the classical mechanics haptic scenarios, and the characteristics of the student participants are described. The evaluation process section defines the experimental and control groups and describes the design and application of the evaluation tools. In the analysis and discussion section, the results are described and assessed. Finally, conclusions and suggestions for future work are provided.

Related work

The potential of using haptic devices to enhance learning is based on the theory of embodied or grounded cognition. According to this framework, bodily experiences are a key component of cognitive and tactile sensations, which stimulate learners to integrate embodied knowledge into abstract concepts (Shapiro, 2010; Barsalou, 2008; Mahon & Caramazza, 2008; Anderson, 2003; Wilson, 2002). Therefore, knowledge emerges from dynamic interactions between the body and the physical world, where people “have perceptual experiences first to construct multimodal representations in order to be able to mentally simulate what is being presented” (Han & Black, 2011, p. 2282). In other words, allowing students to manipulate experiments, either physically or virtually, helps them learn physics concepts (Zacharia & Olympiou, 2011). Thus, visuo-haptic scenarios may be an effective alternative to experiments in a real laboratory.

Visuo-haptic simulators have been developed for different areas and levels of physics learning (from elementary to undergraduate courses) (Okamura, Richard, & Cutkosky, 2002; Williams, Chen & Seaton, 2003; Williams et al., 2007; Han & Black, 2011; Hamza-Lup & Baird, 2012). For example, for classical mechanics concepts, simulators that help users understand the operation of simple machines (Williams, Chen & Seaton, 2003), gears (Han & Black, 2011), and the nature of friction forces (Hamza-Lup & Baird, 2012), among other topics have been developed. In addition, for concepts related to electromagnetism, virtual environments have been designed to help students comprehend electrical interactions between molecules (Höst et al., 2013) and the strength of electromagnetic forces produced by different electrical charges or electrical current distributions (Sanchez et al., 2013; Neri et al., 2015).

Several studies have been conducted to assess the effectiveness of the implementation of haptic technology on the learning of physics concepts. Some of these studies were based on student perceptions while training with visuo-haptic simulators, while others focused on assessing student learning gains when using such simulators. Generally, most perception studies state that students were satisfied with this kind of experience, and they also affirm that this technology supports the understanding of physics concepts (Williams, Chen & Seaton, 2003; Hamza-Lup & Baird, 2012; Neri et al., 2015). On the other hand, learning gains studies have not shown conclusive evidence that experimental groups obtain greater learning gains than control groups. For example, Han and Black (2011) and Hamza-Lup and Baird (2012) reported greater learning gains for experimental groups, while Sanchez et al. (2013) did not find any differences between experimental and control groups. Similarly, Neri et al. (2015) found greater learning gains for experimental students; however, these gains were not statistically different from those obtained by their control group.
Experiment design

Design process for visuo-haptic scenarios

One of the biggest challenges students face in learning physics is understanding the nature, magnitude, and direction of the different forces exerted on bodies in experimental settings. Gravity, friction, and contact forces applied with different strengths and angles can be confusing and may cause students to misunderstand the laws of physics. Taking these ideas into consideration, we designed appropriate educational resources using visuo-haptic simulators to improve learning physics concepts, specifically the behavior of bodies subject to various forces.

Our first attempts at designing visuo-haptic scenarios focused on an electricity and magnetism course for undergraduate engineering students (Sanchez et al., 2013; Neri et al., 2015). These scenarios were rather simple and did not include sufficient or appropriate visual stimuli; therefore, the scenarios did not take full advantage of the potential of combining virtual scenarios with the perception of feedback forces. In addition, control groups were not considered or were insufficiently defined in those studies to effectively assess the usefulness of the visuo-haptic simulators.

Thus, in this study, a careful process to design learning resources for the participating groups of students was performed using a blended learning framework. Visual and written learning materials for the control group were developed, while a more detailed haptic implementation was developed for the experimental group. In addition, topics related to classical mechanics where a visuo-haptic environment could add value to support learning were identified. Then, a group of expert professors designed scenarios with appropriate variables that could be modified in an interactive manner such that students could feel the corresponding effects on feedback forces. For example, students could modify the values of a block's mass, friction coefficients, or incline angle to improve their understanding of the corresponding physical phenomena. Easy-to-visualize elements were incorporated into the simulators by combining the potential of applying forces through the haptic devices with attractive graphic virtual scenarios in which the simulations obeyed the laws of physics.

Application development

Considering the above design criteria, four visuo-haptic scenarios with haptic feedback were developed to help teach and evaluate classical mechanics concepts. The selected concepts were (a) a block on a rough incline, (b) a rotating door, (c) a double Atwood’s machine, and (d) a rolling reel. The architecture of the developed applications follows that of Salisbury, Conti, and Barbagli (2004). The architecture used to develop the visuo-haptic simulators is shown in Figure 2. In this case, a Novint Falcon haptic device was used to provide force feedback with three degrees of freedom, and the Chai3D framework was employed for graphic and haptic rendering (Conti et al., 2003). The parameters in each simulation were controlled and displayed using the ImGui graphical user interface (GUI) library. Visualizations of each simulator and the corresponding user input data screen are shown in Figures 3 to 6. Each scenario is described in detail in the following subsections.

**Figure 2. Architecture used in the development of the visuo-haptic simulators**

**Scenario A: block on a rough incline**

The main purpose of this scenario was to allow students to feel the friction force $F_{fr}$ when they pushed a block up a rough incline (Figure 3). Here the user manipulated a virtual hand to push the block with applied force $F_{app}$. The controllable parameters of this simulation were the static friction coefficient $\mu_s$, the kinetic friction...
coefficient \( \mu_k \), the block’s mass \( M \), and the angle of the incline \( \theta \). Note that the latter was always set parallel to the incline. The intervals for each parameter were \( \mu_s, \mu_k \in [0.1, 1] \) (for simplicity, the simulator used \( \mu_k = 0.7 \mu_s \)), \( M \in [0.1, 3] \) kg, and \( \theta \in [0, 90]^{\circ} \). The value range was set such that \( F_{\text{app}} \) never exceeded 9.0 N, which is the maximum value that can be displayed by the haptic device. The values of \( \mu_s, M, \) and \( \theta \) could be changed freely to generate new experiments. The force of friction (either static or kinetic) was felt by the user by means of the force applied on the block (\( F_{\text{app}} \)) using the haptic device.

Figure 3. Block on a rough incline scenario

Scenario B: rotating door

The goal of this simulator was to enable the students to recognize the concepts of torque and a lever arm and understand their roles in rotational equilibrium (Figure 4). In this simulation, the door had a fixed width of 1.4 m. The system provided an initial virtual force \( F_v \) produced by a virtual hand that pushed the door, which could rotate freely around a fixed vertical axis. Here, the user could set the strength of \( F_v \), its lever arm value \( r_v \), and the door's moment of inertia around the fixed vertical axis \( I \). The user controlled another hand to apply a force to the door \( F_{\text{app}} \) using a variable lever arm \( r_a \) to keep the door at rest. This experiment was designed to enable the student to find the relationship between \( r_a \) and \( F_{\text{app}} \) to satisfy this goal. Note that, in the simulator, forces \( F_v \) and \( F_{\text{app}} \) were always set perpendicular to the door. If the user did not apply \( F_{\text{app}} \) force, the door would continue rotating around the vertical axis until the user interacted with it. Here, the value ranges of the variables were \( F_v \in [0.1, 9.0] \) N, \( F_{\text{app}} \in [0.1, 9.0] \) N, \( r_v \in [0.1, 1.4] \) m, and \( r_a \in [0.1, 1.4] \) m.

Figure 4. Rotating door scenario

Scenario C: double Atwood’s machine

In this scenario, a double Atwood’s machine was implemented. The simulation showed a double pulley that could freely rotate around a horizontal fixed axis at its center and perpendicular to the plane of the figure (Figure 5). Here, the pulley’s moment of inertia around the axis is denoted \( I \), and a block of mass \( M \) was attached to a rope on the left side of the pulley, which was wrapped around the external radius of the pulley \( R \). A handle was attached to a second rope, which was wrapped around the internal radius of the pulley \( r \). Note that the ropes did
not slip around the pulley at their respective radii. The user had to apply force $F_{\text{app}}$ to pull the handle down in order to set the system in rotational equilibrium. Therefore, this simulator was designed to enable the students to understand the relationships among $F_{\text{app}}$, $M$, $R$, and $r$. If the handle was released, the system would rotate counterclockwise due to the pull of the left mass $M$. The value ranges used in this case were $I \in [0.05, 0.20] \text{ kg} \cdot \text{m}^2$, $M \in [0.05, 2.0] \text{ kg}$, $R \in [0.1, 0.5] \text{ m}$, and $r \in [0.1, R] \text{ m}$.

**Scenario D: rolling reel**

In this virtual environment, a reel located on a horizontal surface was displayed. The reel could roll without slipping and could be pulled using a rope wrapped around an inner pulley on the reel. The system was connected to an ideal pulley on the right side of the scene (Figure 6). The ideal pulley could rotate freely around its fixed axis, and the user was required to pull the rope down to start the motion in the simulation. Here, the objective was to enable the students to understand that, the linear acceleration of the reel’s center of mass will increase when the radius value of the inner pulley $r$ increases for a given downward force $F_{\text{app}}$. Here, acceleration was maximum when $r$ reached the radius of the reel $R$. The reel’s mass $M$ was constant, and, for simplicity, it was assumed that the reel’s moment of inertia was $I = \frac{1}{2}MR^2$. The value ranges of the input variables were $R \in [0.1, 0.5] \text{ m}$, $r \in [0.1, R] \text{ m}$, and $M \in [0.05, 2.0] \text{ kg}$.

**Participants**

During the January to May 2016 term, 85 undergraduate engineering students at Tecnológico de Monterrey, Mexico City Campus enrolled in a first semester physics course were selected and divided into two groups, i.e., experimental (E) and control (C) groups. Each group was formed by two sections of the same course. The same professor taught the two control sections and one experimental section, and another professor taught the second experimental section. The student population $N$ for each group was $N_E = 47$ and $N_C = 38$ for the experimental and control groups, respectively. The experimental group used the visuo-haptic simulators as additional class resources, and the control group used equivalent written materials as part of the normal course lectures.
We ensured that students in both groups had similar backgrounds in physics and academic conditions before starting the comparison study. First, all students in the institution must take the same tests and satisfy the same academic prerequisites to enroll in their first semester courses. Therefore, students of different academic levels are distributed evenly in the different sections of courses; thus, a priori, there is no reason to claim one particular section has better students than any another. On the other hand, the teachers involved in this study can also assert this point because they have more than 20 years of experience teaching this course, and, consequently, they possess general knowledge about the academic profiles of students who typically enroll in their sections. Second, the two teachers involved in the study work in a coordinated manner. For example, they follow the same weekly syllabus, use the same textbook and learning resources, offer similar content and difficulty level, and apply similar learning activities in their corresponding sections. Therefore, it can be affirmed with high confidence that the experimental and control groups had similar backgrounds and education levels prior to the haptic intervention. Nevertheless, the authors recognize the need to assess the academic profiles of students to strengthen this assertion, which will be addressed in future work.

In addition, the students in both the experimental and control groups were exposed to similar content and instruction during this study. The experimental group worked with the visuo-haptic simulators, and the control group was given equivalent written materials covering the same topics.

**Evaluation process**

In this section, the testing process and assessment tools used in this study are described.

**Implementation process**

The students in the experimental group worked in pairs to solve a set of carefully designed activities using the visuo-haptic scenarios during two two-hour sessions. In the first session, students worked with Scenario A, and, in the second session, the students worked with Scenarios B, C, and D. At the beginning of each session, 10 to 15 minutes were provided as a free exploration period to allow the students to familiarize themselves with the haptic simulator. In the remaining time, the students worked on a series of tasks following a carefully designed *activity guide* that was provided before the students started working with the haptic devices. For example, for the rough incline scenario, the activities included (a) gradually increase the coefficient of static friction $\mu_s$ to determine the applied force $F_{app}$ required to set the block in motion for an initial tilt angle $\theta = 0^\circ$, and (b) for a given initial $\mu_s$ value, gradually increase tilt angle $\theta$ and study the relationship between $F_{app}$ and the resulting frictional forces $F_t$ acting on the block.

At the end of each session, the students were asked to individually complete a perception questionnaire (PQ) about their experience with the haptic simulators. In addition, each pair of students was required to deliver a short report about the experiments performed during the session.

**Assessment tools**

*Perception questionnaire*

The PQ was applied at the end of each session and focused on the students’ experiences while working with Scenario A (first session) and Scenarios B, C, and D (second session). Students were asked to choose which scenario they liked the most and which one they liked the least. The questions are shown in Table 1, where questions Q1 to Q6 were answered in both sessions. In the second session, Q7 and Q8 were added as general experience questions. Note that the PQ was based on a five-point Likert scale for questions Q1 to Q6. In addition, the students were asked to provide free comments about their experience working with the visuo-haptic simulators.

<table>
<thead>
<tr>
<th>Question number</th>
<th>Statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>Please rate the difficulty level to select the parameters of the simulation.</td>
</tr>
<tr>
<td>Q2</td>
<td>Please rate the difficulty level for using the haptic device in the simulation.</td>
</tr>
<tr>
<td>Q3</td>
<td>How comfortable did you feel yourself when visualizing the 3D environment?</td>
</tr>
<tr>
<td>Q4</td>
<td>How realistic was the visual perception of objects in the 3D environment?</td>
</tr>
</tbody>
</table>

*Table 1. Perception questionnaire*
Q5 How realistic was the tactile perception of objects in the simulation?
Q6 Do you consider that the use of visuo-haptic simulators supports your learning process of course concepts?
Q7 Which one of the 4 experiments did you like the most? Why?
Q8 Which one of the 4 experiments did you like the least? Why?

Post-tests

After the test period, students from both the experimental and control groups were asked to answer two post-tests in class, one for each session. The post-tests covered concepts practiced in the visuo-haptic experiments. The primary goal of these tests was to assess and compare possible learning gains between each group. Post-test one was designed for Scenario A and comprised four questions, and post-test two was designed for Scenarios B, C, and D and included five questions. The students were allowed 25 minutes to answer each post-test, where they were asked to explain each answer clearly.

Results and discussion

In this section, we provide a summary and analysis of the results obtained from the PQs and post-tests.

Perception questionnaires

The results of questions Q1 to Q6 of the PQs are shown in Figures 7 to 12, respectively, where (a) refers to Scenario A and (b) to Scenarios B, C, and D. Here, the vertical axes indicate the frequency of students who answered the corresponding question, and the horizontal axes represent the five-point Likert scale. The steps of the Likert scale are defined in each figure.

![Figure 7](image1.png)

**Figure 7.** Question Q1: difficulty level to select the simulation parameters (1 = very difficult, 2 = difficult, 3 = neutral, 4 = easy, 5 = very easy)

![Figure 8](image2.png)

**Figure 8.** Question Q2: difficulty level using the haptic devices in the simulation (1 = very difficult, 2 = difficult, 3 = neutral, 4 = easy, 5 = very easy)
Figure 9. Question Q3: degree of student comfort visualizing the 3D environment (1 = very uncomfortable, 2 = uncomfortable, 3 = neutral, 4 = comfortable, 5 = very comfortable)

Figure 10. Question Q4: realism of visual perception of objects (1 = totally unrealistic, 2 = unrealistic, 3 = neutral, 4 = realistic, 5 = totally realistic)

Figure 11. Question Q5: realism of tactile perception of objects (1 = totally unrealistic, 2 = unrealistic, 3 = neutral, 4 = realistic, 5 = totally realistic)

Figure 12. Question Q6: visuo-haptic simulators support the learning of course concepts (1 = total disagreement, 2 = disagreement, 3 = neutral, 4 = agreement, 5 = total agreement)
Figures 7 to 12 show that most students generally considered that (a) the visuo-haptic simulators were reasonably easy to manipulate, (b) the visualization of objects was adequate, (c) the visual and tactile realism was appropriate, (d) the information provided was useful, and (e) the simulators were suitable to support class lectures. Nevertheless, in all cases, the perceived perception was better for Scenario A than for the other three scenarios. This result may reflect that students were already familiar with problems that deal with inclines, the design of Scenario A was comparatively more realistic, intuitive, and easier to handle, or that students had more time to practice Scenario A compared to the other scenarios.

**Figure 13.** Question Q7: which of the four experiments did you like the most?

The results for questions Q7 and Q8 are shown in Figures 13 and 14, respectively. These pie charts show that Scenario A was the most liked (53.1%). In addition, it is interesting to note a dual perception relative to Scenario B. In Q7, Scenario B was the second most liked (28.6%); however, in Q8, Scenario B was the least liked (36.7%). For Scenario D, the overall results indicate that it was the least liked (Q7: 6.1%; Q8: 26.5%). Consequently, we consider that Scenario D must be redesigned for future implementations.

**Figure 14.** Question Q8: which of the four experiments did you like the least?

**Student comments**

The student comments regarding the simulators are summarized as follows.

- For Scenario A, most students stated that it was the easiest to understand, was the most realistic, and was the most interactive, dynamic, and easy to control. Therefore, the underlying theory was easier to grasp. However, some students stated that the visualization required improvement and they required more time to work on it.
- Scenario B possessed more variables to manipulate. It was exciting to push the door and interact against the virtual force. Scenario B provided a 3D visualization and can be considered an everyday activity. In addition, the guide for this exercise required less data to be written down. On the other hand, some students said that the simulator did not have clear instructions, and they agreed that it was difficult to control the hand to provide $F_{app}$ and visualize both hands while the door was rotating.
Most students found Scenario C realistic and clarifying, and it helped them learn the related concepts. They considered it the simplest scenario in terms of visualization and control. As with the rotating door scenario, some students expressed that the instructions were not sufficiently clear. Similarly, several students commented that it was somewhat tedious to practice this scenario and that changing the physical parameters on the screen was not sufficiently intuitive or user friendly.

Finally, most students stated that Scenario D helped them understand the concepts, and they expressed that “it was kind of cool.” Nevertheless, most students found that it was difficult to obtain data in this scenario. They had problems finding the lever (3D spatial location) required to pull the reel, and they found it difficult to apply constant force in order to accelerate the reel. Consequently, Scenario D was considered the most complicated scenario.

Post-tests

The results of the post-tests for the four scenarios are presented in Table 2, where E and C represent the experimental and control groups, respectively. The student populations for these groups were \( N_E = 47 \) and \( N_C = 38 \), respectively. Note that the grading scale was normalized to \([0, 100]\) for all scenarios.

As can be seen in Table 2, the post-test averages are better for the experimental group than for the control group for Scenarios A, B, and C. However, for Scenario D, the control group obtained slightly higher average results than the experimental group. Here, \( t \)-tests were performed to assess the statistical significance of these differences using the following null hypothesis.

\[
H_0: \langle \text{Post-test} \rangle_{\text{EXPERIMENTAL}} = \langle \text{Post-test} \rangle_{\text{CONTROL}}
\]

where \( \langle \text{Post-test} \rangle \) is the post-test average. The \( p \)-values for each scenario are shown in Table 3.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Group</th>
<th>Post-test average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Block on incline</td>
<td>E</td>
<td>58.1</td>
<td>17.8</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>37.0</td>
<td>12.9</td>
</tr>
<tr>
<td>B. Rotating door</td>
<td>E</td>
<td>70.2</td>
<td>24.2</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>63.5</td>
<td>23.6</td>
</tr>
<tr>
<td>C. Double Atwood's machine</td>
<td>E</td>
<td>70.2</td>
<td>26.5</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>56.6</td>
<td>20.1</td>
</tr>
<tr>
<td>D. Rolling reel</td>
<td>E</td>
<td>31.4</td>
<td>18.3</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>32.9</td>
<td>15.2</td>
</tr>
</tbody>
</table>

Table 3. \( p \)-values of each visuo-haptic scenario

<table>
<thead>
<tr>
<th>A. Block on an incline</th>
<th>B. Rotating door</th>
<th>C. Atwood's machine</th>
<th>D. Rolling reel</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p )-value</td>
<td>1.679 ×10^{-6}</td>
<td>0.1921</td>
<td>0.03419</td>
</tr>
</tbody>
</table>

The null hypothesis was rejected for Scenarios A and C, which implies that the experimental group obtained a better statistically significant post-test average than the control group for these scenarios. This difference is particularly noteworthy for Scenario A. Generally, this result suggests that the carefully designed visuo-haptic scenarios promoted greater student learning gains than traditional instruction methods. On the other hand, the differences for Scenarios B and D were not significant, which indicates it is necessary to improve their design (particularly Scenario D) in order to take better advantage of haptic devices. Moreover, these scenarios can be improved by designing additional tasks and exercises to better illustrate the relationships among force, the lever arm, torque, and linear and/or angular acceleration.

In summary, from the PQ and post-test results, we assert the following.

- Scenario A was well designed. It is reasonably easy to visualize, handle, and understand. Therefore, this scenario was the most liked. Through the sense of touch, students were able to feel and compare the static and kinetic friction forces. Note that students in the experimental group obtained statistically significant greater learning gains than the control group \( (p = 2 \times 10^{-6}) \). These results concur with those of Hamza-Lup and Baird (2012). Here, we did not attempt to design a similar real experimental setup in the lab and compare it to the visuo-haptic simulator, as Hamza-Lup and Baird did, because the main objective of this study was to allow the students to easily change the physical parameters of the haptic environment. This can
be achieved in a straightforward manner using a virtual simulator; however, in a real lab, it would require implementing different settings for different sets of variables.

- Although the general student perceptions for Scenarios B, C, and D were positive, some students revealed certain difficulties while interacting with the simulators. The manipulation of the objects and selection of the physical parameters must be improved to make them more engaging and user-friendly. On the other hand, significant learning gain differences between the experimental and control groups were only obtained for Scenario C ($p \approx 0.03$). For Scenario B, there was a better learning gain average for the experimental group; however, the difference with the control group was not statistically significant. With Scenario D, there was no difference between the groups. Therefore, to obtain better results, it is necessary to implement three separate working sessions, one for each simulator, in addition to the design improvements mentioned above. This would allow the students to become more familiar with each environment prior to interacting with the environments. The authors also consider that the number of post-test questions used to assess the learning impact of these scenarios should be increased to obtain a more detailed comparison of the learning gain differences between the experimental and control groups. Finally, as mentioned above for scenario A, relative to the scenarios presented in this research, visuo-haptic simulators can be considered viable learning alternatives to real physical models.

Conclusion and future work

The design of visuo-haptic scenarios requires the interaction of different areas of knowledge that combine efforts and skills from educators and technology developers to create comprehensive resources aimed at achieving meaningful learning. Such resources must incorporate the sense of touch to help students feel and handle objects in a learning environment in a straightforward manner, making visuo-haptic simulators an additional and powerful educational tool. Therefore, incorporating haptic technology in learning simulators to teach physics has great potential, and it can enhance the understanding of physics concepts.

In this study, we designed and tested four visuo-haptic simulators for classical mechanics to assess their effectiveness in student learning. For this purpose, an architecture and process to facilitate the development of visuo-haptic applications to teach physics concepts were defined. Using haptic technology, tactile sensations could stimulate the learners’ integration of embodied knowledge in their cognitive processes while learning abstract scientific concepts. Engineering students from a first semester physics course at Tecnológico de Monterrey, Mexico City Campus practiced with the visuo-haptic simulators. The implementation of guided activities for each scenario allowed the students to first develop a hypothesis and then test it using the haptic simulators. The visual cues and aids implemented in the environments helped the students build connections between the data presented in the GUI, the perceived force feedback, the graphical representation of the phenomena, and their theoretical knowledge.

Generally, the results indicate that the students were highly motivated with this technology, and they found it innovative. The challenge for instructors is to take advantage of this fact when designing appropriate haptic interventions that promote better learning outcomes. The results also suggest that visuo-haptic simulators have the potential to improve learning classical mechanics concepts by combining the senses of touch and sight. For Scenarios A, B, and C, the learning gains of the experimental group were greater than that of the control group, although the difference was only statistically significant for Scenarios A and C. However, for Scenario D, no difference was obtained. Therefore, it is necessary to carefully design visuo-haptic simulators to realize more realistic visualizations. Thus, to offer intuitive, friendly, and accurate interactions, we recommend that studies focus on incorporating the sense of depth (3D) and synchronization of forces. The latter can be explained as the way a system handles the force rendered by the haptic device and the motion of objects in the virtual simulator. In addition, we emphasize that it is very important to allow the users to have free exploration time prior to performing the experiments, which will allow them to become familiar with the visuo-haptic simulators.

For future work, we are currently designing further visuo-haptic scenarios to cover additional undergraduate physics topics to expand the study of the impact of haptic technology on student learning. In addition, we intend to perform more extensive studies with greater numbers of students, which will provide stronger constraints relative to the different learning gains of the experimental and control groups. In particular, careful assessment of the similarity of the academic profiles of these groups deserves further attention, which will facilitate deeper insight to better test our hypothesis that visuo-haptic simulators are valuable resources to improve learning physics concepts. Finally, another future objective is to extend the use of visuo-haptic simulators to high school or other academic levels.
References


The Effects of Representation Tool (Visible-Annotation) Types to Support Knowledge Building in Computer-Supported Collaborative Learning

Yoonhee Shin, Dongsik Kim* and Jaewon Jung*

Department of Educational Technology, Hanyang University, Seoul, South Korea // shinyoonhee06@gmail.com // kimdsik@hanyang.ac.kr // jjungi5@gmail.com

*Corresponding authors

(Submitted August 23, 2016; Revised February 1, 2017; Accepted February 4, 2017)

ABSTRACT

This study investigated the effectiveness of visible-annotation types used in online discussions to enhance the accuracy of shared knowledge and the level of constructed knowledge in computer-supported collaborative learning environments. Three types of visible-annotation were developed specifically for this study, which had the function of linking learners’ annotations to related learning content. The participants were randomly assigned into one of three groups. Each group was provided with different types of visible annotations for building shared knowledge in computer-supported collaborative learning: the TL-type for two content-understanding learning phases; the TLL-type for one concept-understanding learning phase and one content-understanding learning phase; and the C-type for one content-understanding learning phase. After completing the knowledge-sharing stage, the accuracy of the participants’ shared knowledge was measured. During the knowledge-construction stage, the participants carried out a lesson-planning task in pairs and constructed knowledge was analyzed based on the lesson plan. The findings revealed that the TLL-type of visible annotation with one concept-understanding learning phase and one content-understanding learning phase was the most effective for enhancing the accuracy of shared knowledge and the level of constructed knowledge in the computer-supported collaborative learning environment.

Keywords

Visible-annotation, Computer-supported collaborative learning, Shared knowledge, Constructed knowledge

Introduction

Collaborative learning is a method that enables learners to share each other’s knowledge and foster collaborative knowledge construction (Eryilmaz, van der Pol, Ryan, Clark, & Mary, 2013). It is effective for enhancing the internalization of knowledge and promoting a higher quality of collaborative knowledge construction (Garrison & Arbaugh, 2007; Morgan, Whorton, & Gunsalus, 2000). In particular, learners use collaborative learning to share knowledge and solve problems in ill-structured problem-solving learning environments (Beers, Kirschner, Boshuizen, & Gijselaers, 2005). Communication activities that include professional knowledge sharing and different perspectives can enhance the level of learning performance (Lomi, Larsen, & Ginsberg, 1997). However, in the process of achieving higher-quality solutions, many restrictions can be imposed due to the learners’ diverse perspectives in such areas as sharing problem awareness, negotiating a variety of opinions, and building collaborative knowledge based on communicative activities (Fuks, Pimantel, & Lucena, 2006). In addition, collaboration load can occur in computer-supported collaborative learning (CSCL) environments, and this can lead to ineffective learning processes and unsuccessful learning performance.

Eryilmaz et al. (2013) argued that collaboration load has consistently occurred in asynchronous discussions. This might have been due to the lack of a shared frame of reference to help team members identify which part of learning content was related to each opinion and individual claim in knowledge building (Hewitt, 2005; Suthers, Vatrapu, Medina, Joseph, & Dwyer, 2008). To build a shared frame of reference in support of online discussions, some annotation functions have been used. The annotation functions provided by “Knowledge forum” and “Concept map” have allowed learners to show all postings in chronological order at one time, write comments pertaining to the related learning content, and share their individual understanding of the learning content. However, when attempts have been made to refer to specific parts of the learning content, the shared frame of reference has not been sufficient to identify the diversity of perspectives and integrate the opinions of team members at one time, which could increase collaboration load.

To overcome these limitations, Eryilmaz et al. (2013) proposed linked annotation. This is an innovative artifact for online discussion in which there is an overarching notion of a challengeable function linking a participant’s contributions with the entire related text provided by the other participants or an instructor. The linked annotation function can play a significant role in forming a shared frame of reference to draw higher-level solutions through meaningful communications. In particular, it effectively supports the building of a common ground and lowers the effort needed to identify the related learning content through minimized coordination. Some empirical studies
exploring the use of linked annotation have shown that collaboration load decreases and interactions such as assertiveness and conflict increase during online discussions (Ding, 2009; Eryilmaz, Alrushiedat, Kasemvilas, Mary, & van der Pol, 2009; Mühlpfordt & Wessner, 2005). Other studies have found that linked annotation has no influence on promoting interactions aimed at clarification and interpretation, and that there is no meaningful effect on collaborative learning outcomes (Eryilmaz et al., 2013; van der Pol, Admiraal, & Simons, 2006). These diverse perspectives might have resulted from a failure to consider the collaborative knowledge-building process when designing a representation tool (Beers et al., 2005; Rummel & Spada, 2005; Slof, Erkens, & Kirschner, 2010; Suther & Hundhausen, 2003).

According to the collaborative knowledge-building process, knowledge-sharing activities that clarify and interpret the meaning of learning content are a prerequisite to knowledge-construction activities that negotiate disparate opinions and derive solutions (Beers et al., 2005; Zhu, 2012). Because the accuracy of shared knowledge has a significant role in building higher-quality constructed knowledge (Barron, 2003; Bromme, 2000; Cannon-Bowers & Salas, 2001; Clark & Brennan, 1991), the learning strategies for building shared knowledge and constructed knowledge should be differentiated. However, in most previous studies, learners have received the same learning methods during both the knowledge-sharing and knowledge-construction stages and the same communication activities have been implemented to build shared and constructed knowledge (Chuy, Zhang, Resendes, Scardamalia & Bereiter, 2011; Eryilmaz et al., 2013; Yücel & Usluel, 2016). Given that the sequential sharing of learning content from the key concept to more complex learning content can effectively enhance the accuracy of shared knowledge and lead to a higher level of constructed knowledge (Beers et al., 2005; Slof et al., 2010), more elaborate learning strategies that consider the collaborative knowledge-building process should be provided to build shared and construed knowledge.

This study was designed to investigate how representation tool types, designated “visible annotation,” enhance the accuracy of shared knowledge and foster the level of constructed knowledge in CSCL environments. Based on the principle of collaborative knowledge construction, we developed the following three types of visible annotations with a linked annotation function: the TL-type, representing the two concept-understanding learning phases necessary to build shared knowledge (TL); the TLL-type representing one concept-understanding learning phase and one content-understanding learning phase used to build shared knowledge (TLL); and the C-type, representing the one content-understanding learning phase used to build shared knowledge (C). The three visible annotation types include the same problem-solving learning phases used to build constructed knowledge.

Specifically, the following research questions were addressed in this study. First, what are the effects of the visible-annotation tool on the accuracy of shared knowledge in the CSCL environment? Second, what are the effects of the visible-annotation tool on the level of constructed knowledge in the CSCL environment?

**Knowledge-building process of CSCL.**

Collaborative learning is effectively carried out if there are active interactions between team members sharing various opinions about learning content (Kolloffel, Eysing, & de Jong, 2011; Suthers & Hundausen, 2003). The process of sharing learning content and problem awareness from diverse perspectives in poorly structured problem-solving learning environments can help learners reduce the possibilities for misunderstanding and lead them to in-depth discussions (Beers et al., 2005; Garrison & Arbaugh, 2007). However, if the number of unshared messages among the learners grows, collaborative interaction is replaced by simple interaction, and this does not positively influence shared knowledge building and collaborative learning outcomes (Lomi et al., 1997; Suthers et al., 2008).

As Beers et al. (2005) and Rummel and Spada (2005) indicated, these limitations might be caused by a failure to consider the collaborative knowledge-building process when designing a representation tool. According to previous research, the limitations on representation tools may be affected if the collaborative knowledge building process is ignored (Slof et al., 2010; Suthers & Hundausen, 2003). Therefore, the appropriate learning processes and learning methods should be applied when carrying out the sequential sharing of learning content from the key concept to more complex learning content (Beers et al., 2005; Cannon-Bowers & Salas, 2001). In addition, the process of collaborative knowledge building should be considered when designing collaborative learning processes to stimulate effective knowledge-sharing activities (Barron, 2003).

Beers et al. (2005) introduced a model that explains the process of collaborative knowledge construction (see Figure 1). The model shows that knowledge shared between team members is constructed through a process of externalization and internalization. New knowledge is then constructed through negotiation and integration.
(Jorczak, 2011; Levesque, Wilson, & Wholey, 2001), which is important to the development of a new understanding based on the consolidation of shared knowledge. In contrast, Rummel and Spada (2005) proposed a collaborative learning model that enables learners to modify and elaborate their understandings through a repetitive process in which different individual perspectives are negotiated. This suggests that the processes of externalization and internalization must be carried out during the entire learning process, not simply before the shared knowledge stage (Ding, 2009; Gao, 2013).

It is therefore apparent that the stages of externalization and internalization are not necessarily prerequisites to the negotiation stage. Accordingly, externalization, internalization, and negotiation could conceivably be applied to each stage simultaneously throughout the entire learning process to build shared and constructed knowledge (Gao, 2013; Rummel & Spada, 2005). In addition, it could be argued that the process of identifying problem awareness and sharing individual understanding is required at the knowledge sharing stage, and furthermore that the negotiation process in which different individual opinions are shared to solve a problem is required at the knowledge construction stage (Beers et al., 2005; Jorczak, 2011; Levesque et al., 2001). The visible annotation developed in this study intends to provide a learning strategy that considers the collaborative knowledge-building process and includes the linked-annotation function to enhance the accuracy of shared knowledge and foster higher-quality knowledge construction for online discussion.

Figure 1. Process of collaborative knowledge construction (Beers et al., 2005, p. 9)

Difficulties of shared knowledge building

Although active interactions are indispensable to effective collaborative learning, it is not easy to derive accurate shared knowledge and higher-quality problem solving from these interactions (Eseryel, Ifenthaler, & Ge, 2013). Even when individuals are provided with the same learning tasks, the degree to which they understand the learning content can be very different. This can result in collaboration load when individuals share their understanding of the content and negotiate various options from different perspectives. Furthermore, these difficulties can expand in a CSCL environment as opposed to a face-to-face environment.

The difficulties of building shared knowledge in a CSCL environment can be explained as follows. First, collaborative learning requires coordinated activities for effective knowledge construction. Although the division of roles, technical support for performing the task, and determination of discussion time are not directly related to the task itself, they are crucial factors leading to fruitful collaborative learning (Gibson, 2001; Kirschner & Erkens, 2013). Therefore, supporting coordinated activities such as collaborative scripts or scaffolding can help learners diminish their collaboration load and engage in both knowledge sharing and construction.

Second, collaborative learning must consider the collaborative knowledge-building process to produce fruitful collaborative learning outcomes. The purpose of collaborative learning is to effectively carry out a complex task (Eryilmaz et al., 2013). However, learners generally experience difficulties with the complex task itself, and this can result in increased collaboration load (Kirschner, Paas, & Kirschner, 2009). In addition, the complexity of the learning content can lead to misunderstandings and result in inaccurate shared knowledge. Therefore, elaborate instructional designs for complex learning are necessary to reduce collaboration load and enhance the accuracy of shared knowledge, as the accuracy of shared knowledge can influence the level of constructed knowledge (Beers et al., 2005). As indicated by Levesque et al. (2001), meaningless interactions can negatively affect collaborative learning. Thus, carrying out shared knowledge-building activities through meaningful interactions based on the collaborative knowledge-building process is required.
Third, the representation tool should provide functions that are appropriate for the learning task (Kolloffel et al., 2011; Suther & Hundhausen, 2003). Given that the representation tool is the only means through which learners with various perspectives can express their opinions in CSCL environments, it should provide proper functions that consider the learning task if it is to lead to successful learning outcomes (Kirschner et al., 2009). A proper, shared frame of reference is essential to foster fruitful knowledge construction and greater gains in collaborative learning outcomes. Functions such as linked annotations and discussion threads support the building of a common ground and lower the effort needed to identify the related learning content. Therefore, it is necessary to design a representation tool that considers not only the collaborative learning process based on complex tasks, but also the distinctive functions that match the features of the learning tasks. The representation tool developed in this study is intended to focus on the collaborative knowledge-building process and the proper functions needed to carry out complex learning, thereby overcoming the difficulties experienced with shared knowledge building.

“Visible annotation” as a more viable representation tool

The development of computer technology has led to a recent interest in representation tools for externalizing individuals’ knowledge within collaborative learning environments. A representation tool can help learners successfully reconstruct learning content and effectively deliver visualized information. In particular, representation functions are required at the troubleshooting stage in the CSCL environment because learners have difficulty sharing opinions and reaching fruitful solutions when they cannot efficiently track what is being discussed (Slof et al., 2010). The representation tool should therefore be supported in the process of learning problem solving skills by presenting annotations together with the related text, sharing opinions, and externalizing knowledge to further successful learning outcomes.

In this study, we developed “visible annotation” as a representation tool of text-based annotations using computer technology. The “linked annotation” function (Eryilmaz et al., 2013) was applied to connect annotations with their related learning content, leading to a shared, cognitive frame of reference and supporting higher quality argumentation. In particular, the tool was very effective in asynchronous discussions because tracking the annotations with their related learning content helped learners to accurately share various opinions and internalize knowledge learned from others. Moreover, the flow of the interactions could be clearly presented. Collaborative learning in problem-solving tasks included sharing learning content, negotiating various opinions, and arriving at solutions. Through this process, visualized information and troubleshooting strategies could be shared between team members and deeper knowledge construction was possible (Jorczak, 2011).

<table>
<thead>
<tr>
<th>Visible annotation type</th>
<th>Knowledge-sharing stage</th>
<th>Knowledge-construction stage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1st week</td>
<td>2nd week</td>
</tr>
<tr>
<td>TL</td>
<td>Content-understanding learning phase for basic understanding of learning content</td>
<td>Content-understanding learning phase for deep understanding of learning content</td>
</tr>
<tr>
<td>TLL</td>
<td>Concept-understanding learning phase for understanding of key concepts</td>
<td>Content-understanding learning phase for understanding of learning content</td>
</tr>
<tr>
<td>C</td>
<td>Content-understanding learning phase for understanding of learning content</td>
<td></td>
</tr>
</tbody>
</table>

The proposed visible annotation was designed by focusing on the more elaborate learning strategies used to build shared knowledge and achieve higher solutions through collaborative learning. It consisted of a knowledge-sharing stage and a knowledge-construction stage. The TLL type of visual annotation used one concept-understanding learning phase and one content-understanding learning phase to build shared knowledge. This was devised to give learners an opportunity to sequentially share learning content from the key concepts to complex learning content. The TL type used two content-understanding learning phases to build shared knowledge. Learners achieved a basic understanding of the learning content in the first phase then deepened their understanding of it in the second phase. In the second phase, learners were instructed to focus on the difficulties that they had experienced in understanding the first phase. The C type used one content-understanding learning phase to build shared knowledge in the same way as the previous representation tools had done. The three types
of visible annotations provided the same problem-solving learning phase to build constructed-knowledge. Tables 1 and 2 show the learning phases and learning methods based on the three different types of visible annotations.

<table>
<thead>
<tr>
<th>Type of learning phase</th>
<th>Learning method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept-understanding</td>
<td>To define the meaning and explain the pros &amp; cons of key terms</td>
</tr>
<tr>
<td>Content-understanding</td>
<td>To ask, explain, and comment on the sentence-based learning content</td>
</tr>
<tr>
<td>Problem-solving</td>
<td>To negotiate various opinions and derive solutions for completing the lesson-planning task</td>
</tr>
</tbody>
</table>

### Methods

#### Participants

Thirty-six students at a four-year college who had enrolled in one or more of four educational technology classes were invited to participate in this study. Sixteen of the students were women and 20 were men with an average age of 21.61. Sixteen were freshmen, 10 were sophomores, and 10 were seniors majoring in Educational Technology. The participants were randomly assigned to one of three groups, and each group was provided with different types of visible annotations for building shared knowledge in a CSCL environment; two content-understanding learning phases (TL); one concept-understanding learning phase and one content-understanding learning phase (TLL); and one content-understanding learning phase (C). The participants in each group performed the learning tasks in pairs. All of the groups carried out a “comprehension of learning content task” and a “lesson planning task” for four weeks within the CSCL environment.

#### Experimental materials

**Pretests**

Five multiple-choice problems were used to measure the participants’ prior knowledge. A five-point Likert scale was developed to measure computer literacy and collaborative preference. Each test consisted of 10 questions (see Table 3). There were no significant differences between the groups in terms of their prior knowledge ($F = 2.226, p = .124$), computer literacy ($F = .660, p = .524$), and collaborative preference ($F = 2.580, p = .091$).

<table>
<thead>
<tr>
<th>Type of pretest</th>
<th>Questions</th>
<th># of items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior knowledge</td>
<td>Choose the wrong explanation about the brainstorming.</td>
<td>5</td>
</tr>
<tr>
<td>Computer literacy</td>
<td>I have no difficulties in learning using the computer.</td>
<td>10</td>
</tr>
<tr>
<td>Collaborative preference</td>
<td>I prefer to exchange opinions with my colleagues.</td>
<td>10</td>
</tr>
</tbody>
</table>

**Measurement**

The independent variables consisted of the three types of visible annotation, and the dependent variables included the accuracy of shared knowledge and level of constructed knowledge. The effects of visible annotation for the enhancement of the shared knowledge were assessed using two computer-based tests, labeling and simple description, after completing the comprehension of learning content task for building shared knowledge. The tests consisted of five labeling problems and five simple pro and con descriptive problems based on previous research by Weinberger, Stegmann, Fischer, and Mandl (2007) (see Table 4). Pair-wise comparison was used and two evaluators were asked to rate the accuracy of shared knowledge on a scale ranging from “inaccurate” (0) in cases where both learners gave the wrong answers to “accurate” (1) in cases where both learners gave the correct answers. If one learner gave a partially correct answer and the other learner gave the correct answer, a partial score of 0.5 was given. Interrater reliability analysis revealed a Cronbach’s alpha value of 0.92.

<table>
<thead>
<tr>
<th>Category</th>
<th>Questions</th>
<th># of items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labeling</td>
<td>An expedition to study something at first hand</td>
<td>5</td>
</tr>
<tr>
<td>Description</td>
<td>To explain the pros and cons of “problem-based learning”</td>
<td>5</td>
</tr>
</tbody>
</table>
Constructed knowledge was analyzed by assessing the lesson-planning task for building constructed knowledge. Learners in each group were required to submit a lesson plan in pairs at the end of CSCL. A three-point Likert scale was developed to measure the level of the constructed knowledge, based on the research of Dick, Carey, and Carey (2004), and Gagné, Wager, Golas, and Keller (2004) (see Table 5). Three evaluators were asked to rate the level of constructed knowledge focusing on the instructional design principle on a scale ranging from “poor” (0) to “excellent” (1). If the lesson plan was rated “fair,” a partial score of 0.5 was given. Interrater reliability analysis revealed a Cronbach’s alpha value of 0.89. Table 6 shows the tasks and measures used to build shared and constructed knowledge.

Table 5. Sample indicators to measure constructed knowledge

<table>
<thead>
<tr>
<th>Category</th>
<th>Questions</th>
<th># of items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame</td>
<td>The learning objective is clearly stated.</td>
<td>8</td>
</tr>
<tr>
<td>Content</td>
<td>Lesson plan offers learning activities appropriate to the learning strategy.</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 6. Tasks and measures to build shared and constructed knowledge

<table>
<thead>
<tr>
<th>Stage</th>
<th>Task</th>
<th>Measure</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge-sharing stage</td>
<td>Comprehension of learning content task</td>
<td>Shared knowledge test</td>
<td>Weinberger et al. (2007)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Two computer-based tests; labeling and simple description after completing comprehension of the learning content task</td>
<td></td>
</tr>
<tr>
<td>Knowledge-construction stage</td>
<td>Lesson-planning task</td>
<td>Constructed knowledge test</td>
<td>Dick et al. (2004)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lesson-planning task based on instructional design principles</td>
<td>Gagné et al. (2004)</td>
</tr>
</tbody>
</table>

Descriptions of visible annotation

The representation tool, visible annotation, was developed by focusing on previous studies of design principles used to develop representation tools and knowledge building theories (http://52.79.61.247). It was based on the Windows 2000 Server operating system. The MySQL Server 2000 was used as a database and the Web server used was CentOs 5.10 of the Windows 2000 Server. Python, HTML, CSS, and JavaScript were used as the programming languages.

The different types of visible annotations with the linked-annotation function consisted of different learning phases and learning methods based on the visible-annotation type used to build shared knowledge and the same learning phase and learning method used to build constructed knowledge (see Table 7).

Table 7. Learning phases and learning methods based on visible annotation types

<table>
<thead>
<tr>
<th>Visible annotation type</th>
<th>Knowledge-sharing stage</th>
<th>Knowledge-construction stage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Comprehension of learning content task</td>
<td>Lesson planning task</td>
</tr>
<tr>
<td></td>
<td>1st week</td>
<td>2nd week</td>
</tr>
<tr>
<td>TL</td>
<td>Content-understanding learning phase</td>
<td>To ask, explain and comment on the sentence-based learning content</td>
</tr>
<tr>
<td></td>
<td>To ask, explain and comment on the sentence-based learning content</td>
<td></td>
</tr>
<tr>
<td>TLL</td>
<td>Concept-understanding learning phase</td>
<td>To define the meaning and explain the pros &amp; cons of key terms</td>
</tr>
<tr>
<td></td>
<td>To define the meaning and explain the pros &amp; cons of key terms</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>Content-understanding learning phase</td>
<td>To ask, explain and comment on the sentence-based learning content</td>
</tr>
</tbody>
</table>

The structure of the visible annotation consisted of a learning phase, learning content, annotation board, and FAQ. Figure 2 illustrates the menus and functions of visible annotation.
The concept-understanding learning phase focused on understanding the key concepts. If learners chose a key word from the learning content section, they could study concept-based knowledge by defining the meaning and explaining the pros and cons of key words. After submitting the concept-based learning task for a key word, annotations could then be presented on the annotation board and the key word in the learning content section could be linked automatically with the annotation. Therefore, if learners clicked on the key word in the learning content section, they could review its definition together with its pros and cons (see Figure 3).

The content-understanding learning phase focused on understanding sentence-based learning content. Learners could study the learning content in this phase by sharing various opinions through questions, explanations, and comments about it. If learners wanted to study specific aspects of the learning content, they could choose a sentence or a paragraph from it. After specifying the range of a sentence or paragraph, they could ask questions, explain what they understood, and provide comments about the specified parts. Annotations could be presented on the annotation board after submitting the content-based learning task. If learners clicked a specific sentence or paragraph from the learning content, it could be linked automatically with the annotation on the annotation board (see Figure 4).
The problem-solving learning phase focused on the lesson planning task at the knowledge-construction stage. Learners could share and negotiate various opinions through questions, explanations, and comments to derive solutions to complete the lesson-planning task. Other functions were provided in the same way as in the content-understanding learning phase (see Figure 5).

**Procedure**

The participants were divided into three groups, with each group being furnished with different types of visible annotations to perform their task. The experiments were conducted in the form of an assignment that lasted four weeks. The participants were first provided with a visible annotation of the CSCL environment and then participated in an experiment in pairs. An instructor provided the participants with special instructions for completing the task without having face-to-face contact.

**Knowledge-sharing stage**

All of the participants were provided with the same learning content using visible annotation in the CSCL environment. The learners carried out the comprehension of learning content task for shared-knowledge building. The learning content was presented as an online text related to learning theories, learning strategies, and instructional design principles. To understand the learning content, the TL group received two content-understanding learning phases. In pairs the learners studied the learning content by sharing their opinions through questions, explanations, and comments. The TLL group received two learning phases, one concept-understanding learning phase and one content-understanding learning phase. In the first phase, the paired learners studied the key concepts, defined the meanings and explained the pros and cons of the key terms in the learning content. In the second phase, they studied the learning content by sharing their opinions through questions, explanations, and comments about specific sentences or paragraphs in the learning content. The C...
group received one content-understanding learning phase. The learners studied the learning content in pairs, sharing various opinions. All of the learning tasks were conducted non-synchronously for two weeks. After completing the knowledge-sharing activities, each participant took a shared knowledge test on, for example, labeling and providing simple descriptions.

Knowledge-construction stage

The lesson-planning task for building constructed knowledge was conducted synchronously for two weeks. All of the learners engaged in the same problem solving learning phase and they shared and negotiated in pairs to achieve higher-quality solutions when carrying out the lesson-planning task. After the learners completed the knowledge-construction activities, the lesson plans were assessed by measuring the level of constructed knowledge (see Figure 6).

<table>
<thead>
<tr>
<th>Pretests</th>
<th>Knowledge-sharing stage</th>
<th>Shared knowledge test</th>
<th>Knowledge-construction stage</th>
<th>Constructed knowledge test</th>
</tr>
</thead>
</table>

*Figure 6. Learning processes in the CSCL environment*

Results and discussion

The accuracy of the shared knowledge

We investigated the effects of using representation tool types to support knowledge building in the CSCL environment. The TLL group had the highest accuracy of shared knowledge and the C group had the lowest accuracy (see Table 8). The results of analysis of variance (ANOVA) of the shared knowledge found significant differences between the conditions \[ F(2, 15) = 5.543, \ p = .016 \] (see Table 9). A post hoc Tukey test showed the difference between TLL and C \[ p = .015 \]. Although no significant differences were found between the other groups, there were differences between TLL and TL \[ p = .084 \] and between TL and C \[ p = .644 \].

Having accurate shared knowledge is very important because it can influence the quality of constructed knowledge (Cannon-Bowers & Salas, 2001; Dechurch & Mesmer-Magnus, 2010). According to previous studies, when learners are building accurate shared knowledge and integrating new understandings into their current knowledge base, inappropriate representations of the discourse and the complexity of finding their own comments can cause them to experience many difficulties within the CSCL environment (Hweitt, 2005; Simons, 2000; Suthers et al., 2008; Veerman, Andriessen, & Kanselaar, 1999). Visible annotation was intended to overcome these limitations by providing a shared frame of reference that could help learners enhance the accuracy of shared knowledge by indicating which part of the learning content was related to each annotation. Through the linked annotation function applied by visible annotation, visualized information and troubleshooting strategies could be shared between team members, and learners could identify the annotations and their related learning content. In addition, the linked annotation function facilitated the internalization and externalization of the learners’ knowledge by representing the annotation with its related learning content.

TLL was found to contribute most effectively to enhancing the accuracy of shared knowledge. It was designed particularly to focus on building accurate shared knowledge, whereas most previous research has focused on enhancing the level of constructed knowledge in CSCL environments (Castek, Beach, Cotanch, & Scott, 2014; Chuy et al., 2011; Eryilmaz et al., 2013; Gao, 2013; Yücel & Usluel, 2016). Although many studies have found that the accuracy of shared knowledge plays a significant role in building higher-quality constructed knowledge (Barron, 2003; Bromme, 2000; Cannon-Bowers & Salas, 2001; Clark & Brennan, 1991), few efforts have explored the representation tools used to build the shared knowledge that leads to fruitful collaborative learning. TLL provided separate learning phases for concept and content understanding to build accurate shared knowledge based on a collaborative knowledge-building process. During the knowledge-sharing stage, learners could share different points of view on the key concepts and adjust misunderstandings related to learning content to establish a shared understanding. The results of this study showed that the learners’ conflicting points of view declined throughout the sequential sharing process, leading to higher levels of shared knowledge. This resulted from their enhanced understanding of the meanings of things.

TL provided two learning phases for content-understanding through the process of asking, explaining, and commenting on learning content. The intent was for learners to share different points of view on sentence-based
learning content in the first phase then deepen their understanding of the learning content in the second phase. However, the findings revealed that TLL was more effective than TL for enhancing the accuracy of shared knowledge. Previous research had shown that the sequential sharing of learning content from a key concept to more complex learning content could be effective for building accurate shared knowledge (Beers et al., 2005; Slof et al., 2010). Applying this to TLL, it was assumed that the first learning phase for shared knowledge building would influence the construction of an accurate concept understanding, and the second learning phase would influence the enhancement of deeper content understanding. Based on the results of this study, we suggest this assumption was correct and that TLL would be a suitable type of visible annotation for enhancing the accuracy of shared knowledge.

Table 8. Mean and standard deviation of the shared knowledge and the constructed knowledge

<table>
<thead>
<tr>
<th>Visible annotation type</th>
<th>Accuracy of shared knowledge</th>
<th>Level of constructed knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>TL</td>
<td>7.42</td>
<td>.80</td>
</tr>
<tr>
<td>TLL</td>
<td>9.17</td>
<td>.75</td>
</tr>
<tr>
<td>C</td>
<td>6.50</td>
<td>2.23</td>
</tr>
</tbody>
</table>

Note. N = 12. M: Mean. SD: Standard deviation. TLL: Two learning phases for concept-understanding activities and content-understanding activities during the knowledge-sharing stage. TL: Two learning phases for content-understanding activities during the knowledge-sharing stage. C: One learning phase for content-understanding activities only during the knowledge-sharing stage.

Table 9. ANOVA of the shared and constructed knowledge

<table>
<thead>
<tr>
<th>Type of knowledge</th>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shared knowledge</td>
<td>Group-inter</td>
<td>22.69</td>
<td>2</td>
<td>11.35</td>
<td>5.543</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>Group-in</td>
<td>30.71</td>
<td>15</td>
<td>2.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>53.40</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constructed</td>
<td>Group-inter</td>
<td>72.69</td>
<td>2</td>
<td>36.35</td>
<td>9.838</td>
<td>0.002</td>
</tr>
<tr>
<td>knowledge</td>
<td>Group-in</td>
<td>55.42</td>
<td>15</td>
<td>3.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>128.11</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. N = 12. SS: Sum of squares. MS: Mean square.

The level of the constructed knowledge

The TLL group had the highest level of constructed knowledge and the C group had the lowest level (see Table 8). The results of ANOVA of the constructed knowledge found significant differences between conditions \( F(2, 15) = 9.838, p = .002 \) (see Table 9). A post hoc Tukey test showed the difference between TLL and C [\( p = .001 \)]. Although no significant differences were found between the other groups, there were differences between TLL and TL [\( p = .072 \)] and between TL and C [\( p = 0.140 \)]. The results of correlation analysis found significant differences between the shared and constructed knowledge [\( p = .021 \)]. The results implied that the accuracy of the shared knowledge could affect the level of the constructed knowledge (see Table 10).

In this study, TLL was found to be the most effective for enhancing the accuracy of the shared knowledge and the level of the constructed knowledge. These findings were consistent with previous research and provided empirical evidence on how the quality of constructed knowledge is affected by the accuracy of shared knowledge (Barron, 2003; Bromme, 2000). Knowledge-sharing activities that clarify the meaning of learning content are a prerequisite to knowledge-construction activities that negotiate various opinions and derive solutions to achieve a high level of constructed knowledge (Levesque et al., 2001). The results of this study demonstrated that the quality of the collaborative knowledge was affected by the accuracy of the shared knowledge. The findings extend the previous finding that at the knowledge-sharing stage, the sequential sharing of learning content from a basic concept to complex content (considering the collaborative knowledge building process) enhanced the accuracy of the shared knowledge and resulted in a higher quality of constructed knowledge.

The correlation between shared and constructed knowledge was examined on the basis of quantitative correlations and according to collaborative knowledge construction theory (Beers et al., 2005). Significant correlations were found between the shared and constructed knowledge, implying that the accuracy of the shared knowledge could affect the level of the constructed knowledge. These results supported the findings of previous research on how the accuracy of shared knowledge leads to a higher quality of constructed knowledge. The results of this study have overarching implications for instructional design. They suggest that visible annotation
with a concept-understanding learning phase used to facilitate knowledge building in the CSCL environment is significantly correlated with building an accurate shared mental model to understand learning content. In other words, providing a concept-understanding of a learning phase can promote understanding of the task itself, which builds accurate shared knowledge and higher-quality knowledge construction.

Table 10. Correlation between the shared and constructed knowledge

<table>
<thead>
<tr>
<th>Type of knowledge</th>
<th>Shared knowledge</th>
<th>Constructed knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shared knowledge</td>
<td>Pearson correlation coefficient</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>( p )</td>
<td></td>
</tr>
<tr>
<td>Constructed knowledge</td>
<td>Pearson correlation coefficient</td>
<td>.537</td>
</tr>
<tr>
<td></td>
<td>( p )</td>
<td></td>
</tr>
</tbody>
</table>

*Note. N = 12.*

Conclusions and limitations

The representation tool can help learners with various perspectives share their opinions with others, build shared knowledge, and construct collaborative knowledge in CSCL environments. However, despite these advantages, the conventional representation tool has limitations. It cannot help shared knowledge reach a higher-quality cognitive domain because the processes leading to problem awareness, opinion sharing, and collaborative troubleshooting have not been fully considered in the conventional representation tools. In this study, visible annotation, which had the function of linking learners’ annotations with related learning content, was developed to overcome these limitations and enhance the accuracy of shared knowledge and the level of constructed knowledge based on collaborative knowledge construction theory.

We proved that TLL with one concept-understanding learning phase and one content-understanding learning phase during the knowledge-sharing stage was the most effective tool for building shared and constructed knowledge and ensuring fruitful collaborative learning. This study demonstrated that visible annotation with a concept-understanding learning phase for shared knowledge building could positively enhance the accuracy of shared knowledge because the concept-understanding activities could promote comprehension of the key concepts and learners could deepen their content understanding during content-understanding activities. In addition, these findings suggested that the accuracy of shared knowledge could lead to higher-quality constructed knowledge. These findings are consistent with previous studies, which have suggested that the accuracy of shared knowledge is an important factor in enhancing the level of constructed knowledge in the CSCL environment (Beers et al., 2005; Cannon-Bowers & Salas, 2001; Eryilmaz et al., 2013; Levesque et al., 2001; Slof et al., 2010). Based on these findings, it can be concluded that TLL with learning activities, in accordance with each learning phase and based on building knowledge theory, can facilitate the accuracy of shared knowledge and the level of constructed knowledge in CSCL environments.

This study might have been limited by its structure. First, 36 students participated in this study, not a small number. However, the number of participants had to be greater to generalize the findings for further study. Second, the study measured the correlation between shared and constructed knowledge to identify the process of collaborative knowledge construction, effectively acting on the visible annotation tool developed for this study. Future studies should explore the processes of building shared and constructed knowledge in each group to achieve successful collaborative learning outcomes.

Acknowledgements

This work was supported by a National Research Foundation of Korea grant funded by the Korean Government [NRF-2014S1A3A2044609].

References


Modelling and Simulating Electronics Knowledge: Conceptual Understanding and Learning through Active Agency

Adrian Twissell

Aylesbury Grammar School, Aylesbury, United Kingdom // atwissell@gmail.com

(Submitted September 14, 2016; Revised November 29, 2016; Accepted January 25, 2017)

ABSTRACT

Abstract electronics concepts are difficult to develop because the phenomena of interest cannot be readily observed. Visualisation skills support learning about electronics and can be applied at different levels of representation and understanding (observable, symbolic and abstract). Providing learners with opportunities to make transitions between these representations and levels can enhance learning. This case study research explores the use of multiple representations in the development of electronics concepts. It focuses on the close observation of learning in context and draws on the findings from semi-structured interviews. The findings show that applying appropriate modelling and simulating strategies enhance the procedures of learning through the learner’s active agency. Four emergent cognitive profiles, which were shown to characterise learning, describe the different ways participants understood electronics concepts.

Keywords

Visualisation, Learning strategy, Modelling, Simulating, Conceptual change

Introduction

This case study research explores the use of external visual representations in support of learning about abstract electronics concepts, within Secondary Design and Technology (D&T) education (UK). It follows a review of visualisation literature (Twissell, 2014) which indicated that visualisation skills are domain specific, and their application context dependent. Visualisation is interpreted as the use of representations, such as circuit diagrams, during the process of inference making (Larkin & Simon, 1987) and how this is applied in the learners’ developing conceptual understanding.

Abstract concepts are difficult to understand because electronics phenomena cannot be readily observed in the “real world”; only the product of operation can be easily observed. In D&T, because the subject is concerned with practical application and problem solving (Black and Harrison, 1994) rather than solely “theory” and “knowledge,” previous research has focused on the creation of models of knowledge that help the learner develop conceptual understanding through practical application (Kimbell, 1994; Kimbell & Perry, 2001). Procedural knowledge, gained through active engagement, is therefore important for the development of expertise where problem solving requires learners to engage with systems in a practical way (McCormick, 1997; Claxton, Hanson & Lucas, 2014). This research explored the role of procedure and how learners applied this to learning about electronics.

Electronics knowledge can be categorised in terms of its “science.” Researchers suggest all electronics phenomena are interpreted relative to concepts of current and voltage (Metioui & Trudel, 2012). Electronics knowledge also relates to categories of analogue and digital circuit types (Duncan, 1997). These perspectives are useful starting points when exploring learners’ knowledge. However “concepts” are open to interpretation in terms of the electronics phenomena actually conceived. Hiley, Brown and McKenzie Smith (2008) posit a distinction between scientific understanding and practical application. Employee experience has indicated that in addition to conceptual knowledge, procedural knowledge linked with application are essential during problem solving, investigating and analysing (Claxton et al., 2014). Here, I consider electronics knowledge in terms of technology and engineering and consequently develop Metioui and Trudel’s (2012) observation by revealing learners’ perspectives in practice.

Rationale

I aimed to identify whether a difference existed between “traditional” electronics (batteries/bulbs) and electronics based around computer programming (microcontrol) within a practical context. Increasingly electronic functioning in commercially manufactured products is achieved using programming; curricula in the UK have followed this development. I wanted to know more about how traditional and programmed approaches supported learning within the context of practical application, and observe the interplay between the associated...
visual representations. This, it was hoped, would clarify the value of learning based on traditional and programmed approaches and reveal how transitions between the associated representations supported learning.

The conceptual framework focused on the observable, symbolic and abstract levels of representation (Wu et al., 2001) and the interaction between these levels. What was the role of practical experience, for example, and how did this support the interplay between representations and levels? A large body of research describes learners’ misconceptions of electronics knowledge, using experimental methodology (Engelhardt & Beichner, 2004). I wanted to clarify the nature of learners’ knowledge in the context of D&T, as research accounts often present the results of inquiry conducted only in closely controlled conditions. Such research does not always recognise the learning context. Learners’ application was therefore important in revealing the specific nature of understanding. Research on electronics learning in context (Metiou & Trudel, 2012) and procedural learning in D&T (McCormick, 2004) has been noted as scarce. Gentner and Gentner’s (1983) electronics specific models of voltage behaviour are isolated examples of attempts to define learners’ individual ways of thinking about electronics. The interplay between representations (translating/making transitions) was identified as a useful contribution to the work of others.

Literature review

Dual Coding Theory (Paivio, 1986) provides a point of departure for exploring verbal and nonverbal aspects of representation use through learners’ visualisation skills. Visualisation has been described as a process, in that it is “the ability to generate, retain, retrieve, and transform well-structured visual images” (Lohman, 1993, p. 3). It also describes visual imagery as artefact (Hoffler, 2010). In the current study visualisation refers to the cognitive processes associated with visual thinking, image, word and symbol manipulation and transformation. The literature is discussed in terms of information processing and representation.

Processing

The literature suggests that individuals process information differently; often described as learning “style” (Dunn, Beaudry, & Klavas, 2002). However Coffield (2006, p. 23), drawing on evidence for identifiable “learning styles,” describes the concept as “theoretically incoherent and conceptually confused.” I adopt a preference for the term “learning strategy” to acknowledge individuals’ differentiated ways of processing information.

Larkin and Simon (1987) explore the synchronous and sequential approaches to making meaning from visual representations. Synchronous representations (circuit diagrams) provide all of the information necessary to instantaneously infer meaning from the image. Sequential representations (programming code) require the learner to read through the representative code word-by-word. This is slower, but provides explanatory advantages, as Lauria (2015) found, over synchronous representations which necessitate an understanding of the symbols/images used. These are useful factors when considering learners’ use of strategy.

Chen et al. (2011) exploit the use of visualisation strategy in a study grounded in electronics education. Their research combines pictorial imagery with concept models (schematics) and suggests this combination enhances learning because learners can more easily “verify and clarify the existing knowledge” (Chen et al., 2011, p. 269) attached to abstract concepts with familiar referents (pictures) and symbolic representations (schematics), particularly when this reflective stage is built into learning schemes. Supporting research suggests that good existing knowledge is needed to enable representation combination; where combination is not successful, learners often focus on a single concrete representation (Seufert, 2003). Similarly where a representation is unfamiliar learners tend to focus only on surface or concrete features (Seufert, 2003). Arnheim (1970) believed that all thinking can eventually be traced to some form of visualisation. Whilst this position is not widely shared (Gardner, 1984), there is a consensus that synchronous approaches provide benefits to learning about abstract concepts, that are not present in logical-mathematical and verbal strategies (Mathewson, 1999).

Representation

Phenomena can be represented at three levels: observable, symbolic and abstract (Wu et al., 2001). Some representations are tangible artefacts (Hoffler, 2010), such as circuit diagrams. Others are learners’ models
representing the phenomena of interest, such as analogies. I discuss the nature of representations using Wu et al.’s (2001) framework as follows.

**Observable level**

Observations of phenomena are possible because the representation is a physical model, such as circuit prototypes. Observations might include the outcome of circuit function, such as light or sound. Observing a tangible prototype, or its functional outcome, represents a concrete experience which means that it describes specific, rather than general (abstract) phenomena that can be readily perceived by the senses. Physical contact with the components of phenomena are said to enhance learning with visualisations (Kosslyn, Ganis & Thompson, 2001). In some cases researchers refer to surface, rather than concrete features when discussing representations and in doing so ascribe meaning to referential elements that, similarly, can be readily observed (Seufert & Brunken, 2006). A numerical value attached to a circuit symbol is an example of an unambiguous electronics-based referential feature.

**Symbolic level**

At the symbolic level, electronics-based signs and symbols include circuit symbols, diagrams, mathematical symbols, truth tables, pictures and programming code. Learning can be described as “the development of cognitive systems [which] depend on signs and representations as mediators” (Hoffman, 2012, p. 185); this allows the communication of knowledge. To enable engagement and communication with electronics concepts, that knowledge requires an external referent, often provided at the symbolic level. Most symbols (programming code accepted) provide synchronous access to knowledge (i.e., symbols allow the visual elements to be viewed at the same time) and therefore enhance its recall (Chen et al., 2011). In practice this means that symbols accelerate cognitive processing through automation (Kirschner, 2002).

Seufert (2003) suggests that a good understanding of the underlying knowledge is needed to enable inferences beyond surface level features of the symbol. Where symbols are combined forming multiple representations, researchers have found this to be more beneficial to learning (Ainsworth, 2006). To avoid increasing the learner’s cognitive load, representations should aim to: “code” multiple elements of representations as one element, automate visualisation tasks and present information in several ways (Kirschner, 2002).

**Abstract level**

The abstract level refers to knowledge which cannot be readily observed. In practice this relates to internal representations constructed by the learner, from external representations which themselves may be constructions by teachers, or from other sources (Treagust & Duit, 2008). These internal representations, or concepts, have been considered to be “abstract theories” which require “complex metaphor [and] abstract language” to enable their communication (Pule & McCardle, 2010, p. 18).

I refer to internal representations as models of knowledge. Models can represent concepts using specific analogies, providing one way to generate understanding about concepts through comparison of the phenomenon with something not immediately connected with it. The “water-in-a-pipe” analogy is a popular comparison between the behaviour of electricity and water. Analogies can be made with specific metaphors as the representative model. Metaphor has been defined as “seeing, experiencing, or talking about something in terms of something else” (Ritchie, 2013, p. 8). “Flowing” related to electrical current is such a metaphor which embodies the concept of movement. Current has commonly been explained with the use of the flowing water and moving crowd analogies (Gentner & Gentner, 1983). Where metaphor links with physical experiences (e.g., “flowing”), it can be described as a grounded metaphor (Ritchie, 2013).

**Computer programming**

There is little research linking electronics, programming and learning. In the current study programming represents electronics through microcontrol and the programming language BASIC. Program code has the potential to provide a sequentially mediated explanation of electronics-based “events and behaviours” (Paivio, 1986, p. 53); I aimed to clarify whether this supported learning.
Lauria (2015), in research with secondary school age children, has shown that the use of programming code which replicates natural language, in combination with a practical approach to simulating the effects of the program in operation, benefit learning about programming. Using simple on-screen graphics and visual feedback from a physical robot, Lauria’s (2015) research also suggests that cognitive load can be reduced leading to enhanced engagement with the concepts of programming. This approach may support the teaching of electronics, where programming is used to represent complex circuits through the application of programming language and active simulation-based learning (using computer simulation or prototyping board). Further inquiry may reveal how the research from computer science can be applied to teaching electronics and learners’ development of understanding.

**Learning processes**

Constructivism is “a theory of knowledge” which asserts the individual’s role in forming, organising and adapting to their experience of the world (von Glasersfeld, 1989, p. 1). Piaget (1955) concentrated on individual’s assimilation (e.g., recognising patterns in new information and organising/coordinating this into existing schemata) and accommodation (e.g., adapting knowledge during interactions with new situations, i.e., translating from one representation type, such as a circuit diagram, to another such as a program) (Bennett & Dunne, 1994). However the role of social setting is also acknowledged as influencing learners’ development (Vygotsky, 1978). Consequently an important emphasis is placed on the learning context to support the analysis of individual experiences. Developing this framework, Bruner’s (1977) taxonomy describes three learning episodes consisting of: acquisition of new knowledge (complementing existing knowledge), transformation (manipulation, analyse or conversion to another form) and evaluation (plausibility of new knowledge in context).

Kolb and Fry’s (1974) experiential learning, known as Kolb’s Learning Cycle involves the application of four different stages: concrete experience (exposure to new experiences), reflective observation (reflection from different perspectives), abstract conceptualisation (personal theory building) and active experimentation (decision making, problem solving and knowledge testing in new situations). The four stages operate along two dimensions consisting of polarised continuums: concrete/abstract (perceptive) and active/reflective (processing). Thus learners engage with new concrete experiences, reflection leads to abstraction and theorising and experimentation allows testing and problem solving in new situations. Key to all of these processes, and therefore of interest in this study, is how experience is transformed in practice, in relation to teaching electronics.

**Conceptual change**

The field of science education has generated a large body of research into conceptual change (Chen et al., 2013; Treagust & Duit, 2008). This provides an in-practice account of how learners construct knowledge by gradually integrating new concepts into existing understanding (Chen et al., 2013) and therefore how learners adjust schemata through transformations of experience.

Chen et al. (2013) suggest four conditions are necessary for conceptual change: (1) learning material triggers dissatisfaction with existing understandings, (2) new concept visualisations provide intelligibility, (3) plausibility of concept is achieved when visualisation can be matched with theoretical understanding and (4) to overcome longstanding misconceptions, visualisation should be linked with manipulation and exploration opportunities (Chen et al., 2013). This developmental model therefore represents the combination of imagery and experience in modifying learners’ existing understanding and here takes account of learners’ multiple perspectives from different phases of the research.

**Methodology**

This research aimed to describe the different ways learners use external representations to construct abstract electronics concepts. It focused on translations of and transitions between multiple representations, differences between “traditional” and programmed electronics, the role of practical application and learners’ conceptions of knowledge. I draw on a constructivist perspective on learning, which views learning as the consequence of procedure (McCormick, 1997), situated within specific contexts (McCormick, 2004). It was important, therefore, to observe participants during realistic learning tasks during their normal learning. I adopted Siegler’s (2005) Microgenetic approach, which involved the close observation of learning activities, at specific points in the learning trajectory, to reveal the strategies used.
Learners draw on strategies which are time, context and task specific; they are the subject of “variability, choice, and change” (Siegler, 2005, p. 771). Older strategies may be used concurrently with newer strategies, or replaced as learners choose between and change approach. Where progression is impeded due to the lack of a known strategy, learners choose adaptively among those that they do know (Siegler, 2005). A key analytical approach within Overlapping Waves Theory is the focus on learners’ strategies which have been shown to vary among individuals, often within the same type of problem solving task. The research adopts a case study strategy, exploring a specific instance of representation use and aimed to reveal the “exemplary knowledge” therein (Thomas, 2011, p. 211). This is described as the specific study of representation use through the translation of and transition between representations as the unit of analysis.

Design

The research uses a cross-comparative approach (Bryman, 2008) and follows a sequential multi-strategy design (Robson, 2011). It was conducted at a Boy’s selective 11-18 Grammar School in the south of England. Participants were drawn from the researcher’s GCSE D&T: Electronic Products class. Seventeen students participated in the lesson tasks and ten students participated in the interviews. The research was organised around two phases (Table 1). In phase 1 learners engaged with three learning activities during two one-hour lessons. These drew on three key conceptions as follows: that translating and making transitions between representations enhance learning (Ainsworth, 2006), that the act of creating a representation enhances learning processes by combining the generated with the given representations (van Meter & Garner, 2005) and that a qualitative difference exists between engagement with diagrammatic and sentential representations (Larkin & Simon, 1987).

<table>
<thead>
<tr>
<th>Table 1. Phase overview</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase</td>
</tr>
<tr>
<td>Lesson 1:</td>
</tr>
<tr>
<td>Task 1</td>
</tr>
<tr>
<td>(n = 17)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Task 2</td>
</tr>
<tr>
<td>(n = 17)</td>
</tr>
<tr>
<td>Lesson 2:</td>
</tr>
<tr>
<td>Task 3</td>
</tr>
<tr>
<td>(n = 17)</td>
</tr>
<tr>
<td>Phase 2</td>
</tr>
<tr>
<td>(n = 10)</td>
</tr>
</tbody>
</table>

During lessons learners were closely observed by the researcher as teacher/observer and a non-participant observer, following the principles of the Microgenetic approach. Observers discussed activities with learners and audio recorded responses. These were transcribed and analysed using the constant comparison technique (Thomas, 2009). In Task 1 learners matched images and words provided as self-adhesive labels and discussed their reasoning for pairing symbols. In Task 2 learners created a circuit diagram from three representations using the software Circuit Wizard (New Wave Concepts, 2012). The task was linked with recent learning about logic gates. Learners worked on either analogue or digital tasks to provide a “logic of comparison” through “two or more meaningfully contrasting cases” (Bryman, 2008, p. 72). This explored the directional proposition (Yin, 2014) that thinking about analogue and digital electronics would reveal qualitative differences in understanding. In Task 3 (lesson 2), learners translated three representations supporting their creation of a program using the software Picaxe Programming Editor 5 (Revolution Education, 1996–2013). The task required that the program represent the functioning of the electronics shown in the three representations provided. Close observation was achieved using the software CamStudio (2013) to digitally record learners’ on-screen activity during lessons, allowing subsequent analysis.

In Phase 2 semi-structured interviews were conducted with consenting students (n = 10), drawing on the findings from Phase 1. Following Wu et al.’s (2001) approach the interview strategy: (1) avoided mentioning electronics concepts unless raised by the student, (2) any responses deemed to be unclear were questioned further and (3) emerging ideas and meanings were explored further, encouraging student discussion. The interviews were audio recorded, transcribed and analysed as previously described.
Findings and analysis

I draw on Wu et al.’s (2001) levels of representation as an analytical framework for the presentation of findings and their analysis; these are discussed together following the inductive approach taken in this study. I focus on the findings from the program generation task (Task 3, Lesson 2) which relate to the Microgenetic approach, and the development of four cognitive profiles, which emerged from the interviews. Appropriate synonyms represent participants.

Table 2. Event summary

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>If-then statement added</td>
<td>If-then statement added</td>
<td>If-then statement added</td>
<td>Opens Circuit Wizard</td>
<td></td>
<td>Adds “if pin”</td>
</tr>
<tr>
<td>2</td>
<td>“Main program” header file opened</td>
<td>“Digital Dice” file opened</td>
<td>Adds variable</td>
<td>Re-starts &amp; adds output code</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Adds output command</td>
<td>Returns to Picaxe</td>
<td>Adds subprogram name</td>
<td>Returns to Picaxe</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Saves</td>
<td>Returns to “Digital Dice” file</td>
<td>Adds loop command</td>
<td>Types “Main”</td>
<td></td>
<td>Adds subprogram</td>
</tr>
<tr>
<td>5</td>
<td>Syntax error check</td>
<td>Returns to Picaxe</td>
<td>Re-arranges tabs/creates indents</td>
<td>Adds variable</td>
<td></td>
<td>Adds colon</td>
</tr>
<tr>
<td>6</td>
<td>Simulation</td>
<td>Coding annotation</td>
<td>Syntax error message</td>
<td>Simulation</td>
<td></td>
<td>Saves</td>
</tr>
<tr>
<td>7</td>
<td>Saves</td>
<td>Begins subprogram “LEDON”</td>
<td>Changes PIC type</td>
<td>Adjusts simulation speed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Coding annotation</td>
<td>Adds annotation</td>
<td>Moves all codes left</td>
<td>Simulation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Simulation</td>
<td>Saves</td>
<td>Corrects syntax error</td>
<td>Simulation</td>
<td></td>
<td>Coding annotation</td>
</tr>
<tr>
<td>10</td>
<td>Coding annotation</td>
<td>Writes output code</td>
<td>New file-trials loop code</td>
<td>Coding annotation</td>
<td></td>
<td>Completes annotation</td>
</tr>
<tr>
<td>11</td>
<td>Saves</td>
<td>Syntax check</td>
<td>Simulation</td>
<td>Saves</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Prints</td>
<td>Simulation</td>
<td>Implements changes</td>
<td>Prints</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Prints</td>
<td>Errors emerge</td>
<td>Opens Circuit Wizard</td>
<td>Adds subprogram “on1”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Simulation</td>
<td>Changes PIC type</td>
<td>Coding annotation</td>
<td>Coding annotation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Error checking</td>
<td></td>
<td></td>
<td>Prints</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Simulation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Coding annotation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Phase 1: Translation activity (Task 3)

Findings from this phase are presented in Table 2, which summarises a thirty minute on-screen learning task. Analysis focuses on students’ translation of three “traditional” electronics reference representations. In the Analogue group these were a pin-out diagram, circuit diagram and event schedule, and in the Digital group a pin-out diagram, circuit diagram and truth table. Analysis explores the process of creating a program to represent the electronics functioning shown in the reference representations. The events, numbered chronologically, appear linear and tightly organised. In practice there were periods of inactivity or re-writing. Participants following a GCSE Computing course are highlighted, to acknowledge their potential greater confidence with programming.

Five out of six participants began the task by focusing on concrete referents within the circuit diagram provided (a transistor for the Analogue group and a logic gate for the Digital group). In doing this, participants followed the constructivist approach to learning which suggests that learning processes begin with reference to existing knowledge (Bruner, 1977; Kolb & Fry, 1974; Piaget, 1955). Existing knowledge is here encapsulated at the symbolic level of the circuit symbol providing fast, automated access to the underlying concept (Kirschner, 2002).

However Luke (Table 2) applies a different approach by opening the software Circuit Wizard and begins to mimic the reference circuit. This is consistent with Luke’s approach to electronics learning, where he favours a practical strategy when problem solving. This is a useful comparison with the other participants, highlighting Luke’s need to model the circuit in practice, helping to stimulate his programming-based response to the task.

Participants’ ability to translate circuit symbols and understand underlying concepts is demonstrated in the conversion of the central processing component to an “if-then” statement. Between Event 2 and 5 (Table 2) participants thus construct code which connects the circuit input and output. It is noteworthy, in terms of conceptual understanding, that David, Fergus and Sam (Table 2) each use a sub-program to ensure that the program will run continually as otherwise it would respond to an input only once (unlike “traditional” circuits). This reflects these students’ deeper conceptual understanding of the differences between traditional electronics and programming; their thinking is at the abstract level of representation. It is also useful to note David’s approach at the beginning of the task. He searches for a computer file from a previous project to view code which is used in the current program (Event 2 and 4, Table 2). This is consistent with David’s overall learning strategy which values personal discovery where progress is unclear.

From Event 6 participants begin to use the software simulation tool (Table 2). This shows how the model is used to clarify the “plausibility” of the program generated (Bruner, 1977, p. 49). The use of the tool supports Kolb and Fry’s (1974) proposition that active experimentation allows knowledge testing in new situations; simulation provides learners with the agency needed to achieve autonomy and task manipulation to support their learning (Chen et al., 2013).

Two outcomes from Phase 1 supported the design and analysis of interview questions. Firstly the use of Circuit Wizard to virtually construct the circuit, thereby providing another opportunity to model and simulate circuit behaviour by some of the participants, led to a question focus on practical strategy when developing understanding. Secondly, the use of more advanced programming procedure, shown in the sub-program, stimulated a focus on the potential for programming to provide a more explanatory representation for traditional circuits during the interviews and their analysis.

Conceptual change and programming

Drawing from Chen et al. (2013), it can be shown that the programming task satisfies the conditions for conceptual change. Taking the four points in turn, programming used to represent traditional circuit functioning provides: (1) the necessary trigger of dissatisfaction exposed by the need to create alternative ways to represent electronics concepts, (2) intelligibility is provided by the new program-based visualisation, (3) plausibility of the analogue/digital concept is achieved when the new program-based visualisation/representation can be matched with theories of voltage behaviour, and (4) manipulation and exploration (learner agency) is provided by opportunities to create the new program-based representation.

Change in conceptual understanding is shown in the use of sub-programs by David, Fergus and Sam. The use of sub-programs reveal an understanding that microcontrol-based electronics are conceptually different than traditional electronics, because the microcontrolled version requires a deliberate coding structure to enable the
constant monitoring of inputs and control of outputs. Conceptual change is therefore inferred through the development of programming understanding which, sequentially, followed learning about traditional electronics.

**Phase 2: Interviews**

The use of cognitive profiles to represent learners’ electronics understandings adapts the method developed by Thomas (2009). Four profiles evolved from the constant comparison technique (Thomas, 2009) and the creation of themes and categories representing learners’ understandings. The profiles represent exemplary cases in each approach to understanding electronics. The profiles developed from responses to questions which explored: how external representations were used to construct circuits and programs, differences between analogue and digital electronics, and the role of practical application and programming during learning. Where the profile was influenced by findings from phase 1, this is indicated within the individual profile. The point about which discussion is made in each quotation is italicised.

**The Operative**

The Operative profile was developed from interview responses indicating that the participant’s conceptual understanding was grounded in observable features of representations and physical contact with tangible features of phenomena. Luke focused on surface elements of representations, for example, he explains his reasoning in relation to component placement:

“Looking at previous circuits so then most of the ones I’ve seen the capacitor’s around here, so I thought it might work there and the same with the LED, it’s always … most of them are on the far right”

The emphasis on spatial location to identify components was unique to Luke and reflects his reliance on this as a problem solving strategy. He relied on existing knowledge to explain understanding and was unable to make transitions between representations. Luke’s understanding is grounded in a practical approach to tasks and tangible features of circuits and functioning. Luke describes electronics as:

“In relation to programming, Luke suggests that there is a qualitative difference between creating a circuit and writing a program, he explained “one’s more practical … and one’s more thinking about it.” The Operative profile therefore describes a learner who translates representations only on the basis of what is already known and information readily understood from surface features. Transition between representations is difficult because the analogical metaphors have not been developed which would support conceptual change. Conceptual understanding remains at the observable level.

**The Logician**

The Logician’s understanding is grounded in logic-based electronics (logic gates/digital circuits). Understanding draws on the principles of digital voltage. Ethan’s (Analogue task) approach to matching the images and words in Task 1 was based on making associations between programming code and digital electronics. He made frequent reference to programming terms such as “high” and “low,” and digital terms such as “pulse,” “0” and “1” when explaining voltage behaviour. He could distinguish between analogue and digital concepts, but the digital approach was more clearly defined, for example:

“… this chip [logic gate] would generate more of a pulse I think so the LED would switch on/off, on/off [digital concept] and this one [RC network] is a sort of time delay. When you switch the switch the capacitor fills and … that causes a time delay”

Logicians make good use of the symbols embedded in representations, focusing on the central process components. Ethan explains:
“… I’d say the transistor’s the one that really defines what the circuit does and what it is … I would sort of visualise what the PIC [Programmable Interface Controller] chip would do if it was trying to complete the same function in this circuit”

The logical approach is also reflected in Ethan’s explanation of the sequential nature of program code and its ability to explain circuit function, as it “… helps you understand the processes in the circuit and what it’s actually doing.” He clarifies:

“With the commands you can then see what they’re doing in the circuit … [it’s] a more visual way of seeing what each of those code lines actually is”

Conceptual understanding revolves around the symbolic and abstract levels. Ethan’s deeper understanding is reflected in his belief that programming takes it “to another level” and reflects his experience of conceptual change - from traditional to program-based electronics.

The Programmer

Whilst Logicians had a good understanding of programming, Programmers evidenced their understanding of electronics almost exclusively through the procedures of programming. Programmers, unsurprisingly, also tended to focus on the digital aspects of voltage behaviour. Connor (Digital task) explains that his use of programming:

“… helped me understand … it’s just an easy way to relate something that sometimes I may find complicated to something that I find relatively easy”

Similarly Feidhlim explains:

“… you can see what the program does, so you can see that when you say high C.2 … it puts the second output high and that makes the LED … go high”

Programmers made use of the explanatory nature of program code to support their understanding and ability to translate representations. The extra time needed to access this sequential representation was acknowledged, however, and the circuit diagram was valued for its ability to provide synchronous access to information, once understood.

The Dialectic

“Dialectic” is taken from the term meaning to understand both sides of an argument. The term most closely represents participants’ understanding of both analogue and digital circuits; in other words Dialectics have an understanding of voltage which is more advanced than those whose understanding is based only on physical contact or logic. Fergus (Analogue task) explains his understanding as:

“An analogue [circuit] I would categorise as the opposite of digital, so it’s more … multiple states, not just two” … “It’s got a number of voltages and currents involved”

Fergus made very few references to practical experience. His understanding was articulated at an abstract level and this understanding supported his efficient translation of representations and ease of transition between them. Dialectics paralleled the other profile types in appreciating the synchronous nature of diagrams to provide immediate information and the ability of program code to explain circuit function.

Discussion

Learning procedure

The close observation of learning revealed a common strategy when problem solving. Strategy involved a focus on concrete aspects of representations and links with existing knowledge. This parallels constructivist theories of learning which suggest that knowledge is gradually and actively constructed, where understanding develops
from earlier schemas which are subject to change (Piaget, 1955; Bruner, 1977). Active agency, provided by ICT-based simulation tools which have been shown to enhance learning (Chen, 2011), was achieved where learners were able to model and reflect on experience, thereby progressing thinking along the concrete-abstract continuum (Kolb & Fry, 1974). Actively constructing a representation was shown to support van Meter and Garner’s (2005) observation that deeper understanding can be achieved when the given and created representations are considered together. Therefore through active agency and the procedures of representation translation and transition, meaning was shown to emerge as a product of the task (McCormick, 1997).

Differentiated knowledge

Meaning was represented by four constructs indicating that knowledge was personal and contextualised. This posits an alternative to Gentner and Gentner’s (1983) moving crowd and water-in-a-pipe conceptions of electronics knowledge. Drawing from Solsona et al.’s (2003) similar profile concept, I named the profiles the Operative, Logician, Programmer and Dialectic. The constructs show that learners develop understanding at the different levels of knowledge representation discussed by Wu et al. (2001) and supports the distinction between scientific knowledge and that revealed through practical application (Hiley, Brown, & McKenzie Smith, 2008).

The Operative creates meaning from what can be recognised in a representation; it’s concrete elements. Physical contact with tangible elements of phenomena is also important. The Logician understands digital electronics, logic and their symbols of representation. Although more abstract in nature than the Operative’s understanding, meaning is grounded in knowledge about the behaviour of digital voltage. The representations used are mainly synchronous. The Programmer creates meaning from sequentially represented program code and its functionality within microcontrolled electronics. Programmers also ground their understanding in digital voltage, but knowledge was communicated through a preference for programmed strategies. The Dialectic grounds understanding in the concepts of analogue and digital voltages. Meaning is conveyed using comparisons between these conceptions, drawing on metaphorical representations, rather than concrete observations “in the real world.”

Conceptual change

Drawing on a model of conceptual change (Chen et al., 2013; Treagust & Duit, 2008), the findings show how learners developed their schemas of understanding through active transformations of experience. New knowledge, in the form of electronics functioning represented as program code, has been gradually integrated with existing understandings grounded in “traditional” electronics. Developing the use of multiple representations has been shown to support the position that representing information in several ways enhances learning (Kirschner, 2002; Siegler, 2005). Developing opportunities for conceptual change, therefore, provides an in-practice strategy to support transformations of experience and learners’ knowledge and understanding.

The four cognitive profiles reflect the specific ways learners created conceptual understandings within D&T electronics. Each profile reflects the learner’s personal transformation of experience. However observations of conceptual change are inferred on the basis that learners began with limited knowledge of electronics. Conceptual change related to programming relies on the assumption that programming experience was limited in relation to its use as a functional representation for electronics knowledge; a reasonable assumption to make in this context.

Conclusion

This case study research shows how external visual representations were used to develop learners’ conceptual understanding of electronics. Findings show a common process of representation use during problem solving and a differentiated approach to concept development. Four cognitive profiles were developed as an electronics specific model of understanding: the Operative, Logician, Programmer and Dialectic, representing the different ways learners created understanding through actively engaging with traditional electronic circuit components and making transitions between these based on programmed methods of representing electronics. Physical engagement with the elements of electronics in the form of modelling and simulation was therefore shown to be an important strategy in learning about complex abstract concepts.
Transformations between traditional and programmed methods of representing concepts were shown to enhance learning in relation to the concept of analogue and digital voltage behaviour. Programming code was shown to provide a sequential and explanatory representation of electronics which supported circuit diagrams and simulated models. Therefore opportunities to engage learners in actively modelling and simulating phenomena, as a visualisation strategy, should be encouraged to support transformations of experience leading to developed meaning and understanding.

The research drew on Siegler’s (2005) Overlapping Waves Theory which suggests that learners use different adaptive strategies at different times during their learning. This was shown to involve the combination of external visual representations, existing mental constructs and the active participation in learning activities. Siegler’s (2005) microgenetic approach was used to guide the close observation of learning in context, providing a snapshot of learners engaged in practical learning tasks and the strategies employed by them at a single point in time. Further research is needed to identify if and how different strategies are used at different points during a learning trajectory. Similarly the findings from semi-structured interviews reflect learners’ perspectives at the time of interview. Adopting a longitudinal approach may show how perspectives change over time and reveal the influences on such change.

References


Representations of Animal Companions on Student Learning Perception: Static, Animated and Tangible

Zhi-Hong Chen\textsuperscript{1}\textsuperscript{*} and Sheng-Chun Wang\textsuperscript{2}

1Graduate Institute of Information and Computer Education, National Taiwan Normal University, Taiwan // 2Department of Information Communication, Yuan Ze University, Taiwan // zhchen@ntnu.edu.tw // s1016431@mail.yzu.edu.tw

\textsuperscript{*}Corresponding author

(Submitted September 25, 2016; Accepted January 12, 2017)

ABSTRACT

Educational agents (including animal companions) are a kind of interface agent, which is a significant component in educational systems, and has great impacts on student learning. However, with the development of emerging technologies (e.g., tangible technology), few studies have yet investigated how tangible technology can be used to combine with animal companions to foster student learning and interaction. In addition, different representation styles of animal companions on student learning are also seldom investigated. To address these research questions, this paper develops a My-Dragon system which blends tangible and animal companions for maximizing young student interaction. The system is further evaluated by an empirical study by eighteen pupils compared with a static and an animated version. The results reveal that such a tangible system is of benefit to the student learning perception in terms of attention, emotion, and interaction aspects. Some implications on findings and future system development are also discussed.

Keywords
Tangible technology, Educational agents, Animal companions, Human-computer interaction, Children’s perception

Introduction

Interface has been regarded as having a crucial influence on the quality of the human-computer interaction (Preece, Rogers, & Sharp, 2002). Because of such significance, the concept of the “interface agent” has been incorporated to foster human-computer interaction, as well as to facilitate the creation of a learning environment, where the interface not only acts as a channel affecting student participation, but also has a critical role on their learning quality. Recently, the development of a learning environment with interface agents is continuously influenced by the two research sub-fields: embodiment technology and educational agents.

Regarding embodiment technology, the human-computer interface has developed from textural to graphical, multimedia, gestured, and tangible mediums, where advanced technology has great impact on the interface design. Among them, tangible technology involves how users interact with digital information by the physical environment (Ishii, 2008), and has been devoted to the application settings, where students use their hands to manipulate physical objects that can detect and give appropriate feedback (Ullmer & Ishii, 2001; Ishii & Ullmer, 1997). For education application, tangible applications might benefit student learning in terms of learning experience and pedagogical strategy. First, tangible technology can transform information from the virtual world to the physical world, which can help students learn through actual operation and manipulation (Xu, Read, Sim & McManus, 2009). This model is related with the paradigm of learning-by-doing, and concrete actions could enrich students’ learning experiences. In addition, tangible technology, which emphasizes manipulating physical objects to learn, can be further integrated with multisensory learning, which offer students rich learning experience by multisensory, such as visual, auditory, kinesthetic, and tactile systems. Second, tangible user interface contains the key characteristic of metaphor (Fishkin, 2004), which could offer more cues and affordances by physical features, and further enhances the integration of pedagogical strategies.

Regarding educational agents, educational agents plays as a significant role of fulfilling interactive characters as specific roles, such as tutors, tutees, learning companions (Woolff, 2009). There are many advantages to the use of educational agents in terms of learning experience and pedagogical strategy. For learning experience, educational agents can offer appropriate support similar to that offered by their learning peers (Chan, 1996), which can enrich the learning experience. For pedagogical strategies, as students interact with educational agents, the agents can be empowered by various pedagogical strategies, such as learning by being taught (Heift, 2015; Hernández et al., 2015), learning by co-working (Chan & Chou, 1997; Ogan et al., 2014), learning by disturbing (Aimeur & Frasson, 1996), learning by teaching (Londos, 2015; Sjödén & Gulz, 2015), and learning by caring (Chen, 2012).
Among educational agents, animal companion systems that extend the concept of learning companions (Chan, 1996) to support young students’ interaction, have attracted increasing research attention (Chen & Chen, 2014; Chen et al., 2007). The animal-like companions are attractive for children because the feedback is underpinned by the psychology of emotional attachment to animals and pets (Melson, 2001; Beck & Katcher, 1996). However, although animal companions might have potentials to strengthen student motivation and participation, few studies investigate their impacts of different representational styles (e.g., picture, animated, and tangible) on student learning. In other words, both embodiment technology and educational agents have potentials in enriching learning experience and pedagogical strategies, but the influences of embodiment technology on the design of educational agents is seldom investigated. To address this issue, this study develops an animal companion system with three representational styles (e.g., picture, animated, and tangible) to investigate their influences on student perception. More specifically, the research questions of this study are (1) How can we integrate animal companions with tangible technology as a learning system with tangible representation style? (2) What are student perceptions for different representations of animal companions among static, animated, and tangible styles?

Literature review

Educational agents are a pedagogical representation of social participation through the embodiment of interactive characters (Woolf, 2009; Chou & Chan, 2003), which has attracted many researchers’ attention, because they can offer appropriate social scaffolding (Vygotsky, 1978). The influences of educational agents on student learning can be roughly classified as three facets: cognitive achievement, affective impact, and behavioral interaction.

First, cognitive achievement refers to pursuing the optimal cognitive achievement that students can obtain, which is underpinned by the assumption of Bloom’s two-sigma problem: one-to-one tutoring can improve student achievement better than conventional instructional methods (Bloom, 1984). Used for one-to-one tutoring, the educational agent fills the role of tutor aimed at improving the student’s learning performance. Several such systems have been developed for learning in different domains, such as AutoTutor for Newtonian physics (Graesser et al., 2008), and the SQL-Tutor for query language of databases (Mitrovic, 2003).

Next, the affective impact is concerned with maximizing the attraction of the educational agent on the student’s perception, which is significant because cognitive gains alone do not enough to encourage them to use an educational system (Bull, 2004; Barnard & Sandberg, 1996; Kay, 1995). Several approaches have been applied to enhance the attractiveness of educational agents, for example by giving them human-like personas (McQuiggan & Lester, 2007) or the characteristics of pet relationship (Chen, 2012). The former strategy equips educational agents with the characteristics of empathy (McQuiggan & Lester, 2007) and politeness (Wang et al., 2008) to strengthen their influences on student perception. In the latter strategy, the pet-like relationship makes educational agents in either personal computers (Chen & Chan, 2014) or handheld devices (Liao et al., 2011) more appealing and thus enhances their impacts on student perceptions.

Third, behavioral interaction refers to using educational agents to foster interaction with students via multisensory channels (e.g., visual, auditory, kinesthetic, and tactile). For instance, the importance of visual channels on student preferences and reactions has been emphasized, such as different visual styles and appearances on students’ choice preferences (Gulz & Haake, 2005; Baylor & Kim, 2003; Baylor et al., 2003). Various levels of the behavior realism of educational agents on the students’ responses have also been investigated (Groom et al., 2009). In addition, some studies have highlighted the impacts of using both auditory and tactile channels. For example, Wang and his colleagues (Wang et al., 2013) used tangible learning companions to enhance students’ English conversation via voice recording.

Recently, animal companion systems that extend the concept of learning companions (Chan, 1996) to support children interaction. The research of animal companions has received increasing attention because animal companions could enhance their affective impacts on students—they not only have animal-like appearances to attract students, but are equipped with pet characteristics (i.e., needing students’ feeding and caring) to maintain their relationship (Chen & Chen, 2014). Nevertheless, it is still unclear how animal companions can be harnessed in human-computer interaction by combining affective impacts and embodiment technology. To fill this research gap, the current study is conducted.
Animal companion system: My-Dragon

Subject domain

The subject domain of the animal companion system is basic Chinese characters for elementary students. Different from spelling languages (e.g., English), Chinese is a stroke-based language developed based on a number of graphic-like characters. Each character consists of one or more radicals, and can be further categorized according to its component radical. For instance, the character of 楊 is composed of the component radical 木 on the left, and the radical 易 on the right. Because of this feature, it has been indicated that highlighting the radicals of Chinese characters is a useful pedagogical strategy (Taft, & Chung, 1999), and the component radical plays a significant role in the recognition of Chinese characters (Feldman & Siok, 1997). To this end, the learning goal of the system emphasizes the component radicals of Chinese characters.

Learning flow

The development of animal companion system is underpinned by the theory of emotional attachment to their pets (Melson, 2001), which argues people (especially children) have a natural tendency to build close relationships with their pets (Beck & Katcher, 1996). Based on this close relationship, the presence of pets or animals can provide people social support that is critical to buffering physiological responses to stressful tasks (Allen et al., 1991). Such supporting and companionship might also benefit students when they meet learning difficulties or frustrations. The learning flow used in the animal companion system involves three phases: caring, learning, and challenging, as illustrated in Figure 1.

- Caring phase: the goal of this phase is to deepen attachment with their animal companions, which is realized by allowing students to keep animal companions that need students’ caring. For instance, the students can make their animal companions happy through patting and playing balls with them. By doing so, the students are encouraged to sustain a long-term participation in human-computer interaction.

- Learning phase: the goal of this phase is to offer learning tasks about identifying component radical of Chinese characters. This intention is further realized by allowing students to feed animal companions correct component radicals of Chinese characters. In addition, to convey system feedback, the emotions of animal companions are shown on the computer screen. In other words, in the animal companion system, feeding is equivalent to learning, which requires learning efforts but can be enjoyable. Besides, the computer screen can also present learning materials and learning status (e.g., how many characters the students learned so far and their mastery) to students.

- Challenging phase: the goal of this phase is to improve students’ learning efficiency, which is realized by challenging students by a set of tasks in a limited time. More specifically, students will face a set of multiple-choice questions, whose goal is to evaluate whether students are skilled at Chinese characters. This intention is realized by a set of game levels, in which the students need to obtain the correct answers within a time limit for each game level. To clear a set of game levels, students need to improve their mastery of Chinese characters. This gives the students more chances to improve what they have learned in the previous phase.

Figure 1. Learning flow used in the animal companion system
Three different versions

To address the research question, three different versions of an animal companion system, named My-Dragon, are developed. The three versions have the same subject domain (i.e., Chinese characters), and contain three phases (i.e., caring, learning, and challenging phases). However, the different feature among the three versions is the representational styles, including static, animated, and tangible styles.

- **Static version:** The static version offers picture-based representational style of animal companions, which offers a baseline condition of representation style while compared with other versions. The feedback of animal companions is presented via a set of pictures. For instance, a single “chewing” picture of the My-Dragon will be shown to students while students feed their My-Dragons Chinese characters in the learning phase. Similar, when students pat their My-Dragons in the caring phase, they receive a single “smiling” picture as feedback; students receive a single “frustrated” picture as feedback while they give a wrong answer in the challenging phase.

- **Animated version:** The animated version offers animation-based representational style of animal companions, which offers an ordinary condition of representation style while compared with other versions. Instead of static picture, the feedback of animal companions is presented by animations while students interact with their My-Dragons. In other words, when students feed Chinese characters in the learning phase, pat their My-Dragons in the caring phase, and give a wrong answer in the challenging phase, they will receive “chewing,” “smiling,” and “frustrated” animations as feedback, respectively. The learning flow and screenshots are illustrated in Figure 1.

- **Tangible version:** The tangible version offers tangible-based representational style of animal companions, which offers an enhanced condition of representation style while compared with other versions. In addition to the function offered in the animated version, this version offers a stuffed dragon with an Arduino microprocessor and several sensors (e.g., magnetic sensor, vibration sensor, WebCam, and markers with QR codes) to enhance students’ interaction and participatory experience, as illustrated in Figure 2. While tangible technology is applied to the design of interaction with animal companions, more physical behaviors can be taken into account. For instance, the sensors within the gloves (e.g., magnet and magnetic sensors) can support students’ behavior of patting their My-Dragons. The sensor within the ball (e.g., vibration sensor) can detect students’ shaking and playing with the My-Dragons. In addition, physical behaviors could bring more learning interaction, which is supported by the device of WebCam and radical cards with QR codes. Students need to recognize correct component radical for each Chinese character, and use corresponding component cards to respond to the My-Dragon.

![Figure 2. The tangible version of animal companion system](image)

Method

**Setting and participants**

To have an explorative understanding about the influences of the My-Dragon system on student perception, a study was conducted via comparing with three versions: static, animated, and tangible versions. In other words, the purpose of this study is to compare students’ perceptions among the three versions. After all of the participants used the three versions, they have common use experiences to evaluate and compare them. The participants were 18 fourth-grade elementary students (aged 10 on average), including 10 males and 8 females.
System instruments

The system instruments consist of three versions of the My-Dragon system. All of the learning systems contain the same learning materials: fifty Chinese characters with five categories of components. The major difference among them is the representation style of the animal companions. More specifically, in the static version, learning materials are presented to the students by pictures. This version is regarded as the baseline system for comparison to other systems. The animated version allows students to learn with animated My-Dragon in the learning flow (i.e., caring, learning, and challenging phases). The tangible version consists of all functions of the My-Dragon. Table 1 summarizes the differences among the three versions.

<table>
<thead>
<tr>
<th></th>
<th>Static version</th>
<th>Animated version</th>
<th>Tangible version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Materials are presented</td>
<td>via statistic dragon pictures without interactive behaviors</td>
<td>via an animated dragon with interactive behaviors</td>
<td>via a tangible dragon with sensors and interactive behaviors</td>
</tr>
</tbody>
</table>

Procedure

As illustrated in Figure 3, the following procedures were employed in this study: (1) System introduction: in the beginning, the students were given a brief instruction session about how to use the systems. (2) System use: each student had a 30-minute session to use the three versions, and experienced their differences. To facilitate their observation in different types of system feedback, they were asked to carry out a practical task in each system: learning ten Chinese characters. (3) Data collection: the students were asked to fill out the ranking scale and answer interview questions.

<table>
<thead>
<tr>
<th>System usage (30 minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
</tr>
<tr>
<td>Static version</td>
</tr>
<tr>
<td>Animated version</td>
</tr>
<tr>
<td>Tangible version</td>
</tr>
<tr>
<td>Data Collection</td>
</tr>
</tbody>
</table>

Figure 3. Procedure conducted in the study

Measurement

To quantify the different perceptions of the three versions, a questionnaire developed by the authors of this study is used. As illustrated in Table 2, the questionnaire includes five items: one item in the cognitive dimension, three items in the affective dimension, and one item in the behavioral dimension. For each item, the participants were asked to choose one from the three versions according to their usage experience. In addition, a quick interview was also conducted to collect students’ reasons for their choices.

<table>
<thead>
<tr>
<th>Table 2. Questionnaire items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimensions</td>
</tr>
<tr>
<td>Cognitive</td>
</tr>
<tr>
<td>Affective</td>
</tr>
<tr>
<td>Behavioral</td>
</tr>
<tr>
<td>Aspects</td>
</tr>
<tr>
<td>Helpfulness</td>
</tr>
<tr>
<td>Attention</td>
</tr>
<tr>
<td>Emotion</td>
</tr>
<tr>
<td>Motivation</td>
</tr>
<tr>
<td>Interaction</td>
</tr>
<tr>
<td>Items</td>
</tr>
<tr>
<td>Which version best helps you learn?</td>
</tr>
<tr>
<td>Which version best attracts your attention?</td>
</tr>
<tr>
<td>Which version makes you feel strong emotions?</td>
</tr>
<tr>
<td>Which version makes you feel more motivated to use it?</td>
</tr>
<tr>
<td>Which version offers the most different ways to interact?</td>
</tr>
</tbody>
</table>
Data analysis

The independent variable of the experiment was the three different versions, whereas the dependent variables of the experiment were students’ ranking scores in each aspect. The Chi-square tests were conducted to further validate its significant difference in each aspect. All these analyses were conducted with a Statistical Package for the Social Science (SPSS Windows version 20).

Results

Cognitive aspect: helpfulness

Table 3 illustrates the ranking results of students’ perception in the helpfulness aspect. The result revealed that the tangible version (39%) helped them learn more than the static version (33%) and animated version (28%), but further Chi-square test did not show statistically significant difference. In other words, the finding demonstrated that the three versions had similar influences on their perceived helpfulness. The interviews give some reasons for positive feedback of the tangible version, such as “this system requires some cards to scan the data on them. So, I can see the components of the characters clearly.” (#8-12) and “because I need to take the cards for the given questions, I thus remember them.” (#2-20) On the other hand, because of the lower efficiency, some students expressed negative feedback on the tangible version, including “the static version is easier to use, and I can spend more time in learning.” (#2-25) and “I spent a lot of time in the tangible version.” (#5-2)

| Table 3. Ranking results of the helpfulness aspect |
|---------------------------------|-----------------|-----------------|-----------------|-----------|
|                                 | Static version  | Animated version | Tangible version | χ²        |
| # of students                   | 6 (33%)         | 5 (28%)          | 7 (39%)          | 0.33      |

Affective aspect: attention, emotion, and motivation

Table 4 shows the ranking results of the students’ perception regarding the attention aspect, indicating that most students (72%) felt that the tangible system best attracted their attention among all three systems, which was higher than those in the animated version (22%) and static version (6%), and had statistical difference (χ² = 13.00, p < .01). It implies that the three systems concerning attracting student attention showed an obvious tendency: tangible version > animated version > static version. According to the interviews, one of the reasons for this result was that the tangible system required interesting tasks to do, which will draw their attention. For example, “this tangible version attracts my attention because I need to scan the QR cards using my hands.” (#2-17) and “the tangible version is more exciting because I have to scan these cards one by one.” (#2-11). Another reason for this result was the tangible version is more lifelike with the tangible technology. For instance, “I like keeping pets, and the dragon in this version is like a real one.” (#2-19) and “it is like a true pet.” (#5-7).

| Table 4. Ranking results of the attention aspect |
|---------------------------------|-----------------|-----------------|-----------------|-----------|
|                                 | Static version  | Animated version | Tangible version | χ²        |
| # of students                   | 1 (6%)          | 4 (22%)          | 13 (72%)         | 13.00**   |

Note. **p < .01.

Table 5 shows the ranking results for students’ perception of the emotional aspect, demonstrating that most students (78%) felt that the tangible version aroused their emotions the most, which was higher than those in the animated version (11%) and static version (11%). Further Chi-square test also show a significant difference (χ² = 16.00, p < .01). In other words, the three versions regarding student emotional arousal showed a clear pattern: tangible version > animated version > static version. In the interviews, some reasons for positive feedback included: “The dragon is realistic. It is very nice to learn accompanied by a realistic partner.” (#2-2), “I like this dragon because it is real.” (#2-20), “I can interact with it, touch it, and play with it.” (#2-9)

| Table 5. Ranking results of the emotion aspect |
|---------------------------------|-----------------|-----------------|-----------------|-----------|
|                                 | Static version  | Animated version | Tangible version | χ²        |
| # of students                   | 2 (11%)         | 2 (11%)          | 14 (78%)         | 16.00**   |

Note. **p < .01.

Table 6 shows the ranking results of students’ perceptions regarding the motivation aspect. The result indicated that most students (50%) felt that the tangible version enhanced their motivation the most, which was higher than
those in the static version (28%) and animated version (22%), but further Chi-squire test did not show statistically significant difference. In short, the ranking revealed a similar impact of the three systems on student motivation. From the interviews, some supporting reasons were: “This is because this dragon is real and touchable.” (#2-25), “It is not only a virtual system on the screen, but also an actual entity to touch and play with.” (#8-12), and “It is interactive, and I can touch the dragon.” (#2-9).

<table>
<thead>
<tr>
<th>Table 6. Ranking results of the motivation aspect</th>
</tr>
</thead>
<tbody>
<tr>
<td># of students</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>5 (28%)</td>
</tr>
</tbody>
</table>

Behavioral aspect: interaction

Table 7 illustrates the ranking results of students’ perceptions of the interaction aspect. All of the students (100%) agreed that the tangible version offered the most engaging way to interact. In other words, the three versions concerning interaction showed a clear tendency: tangible version > animated version or static version. According to the interviews, some positive feedback of the tangible version included: “in addition to playing and patting, I need to move it to other areas for different tasks.” (#2-24), “I can use a ball to play with it, touch it with a glove, and communicate with it by cards.” (#5-7), and “I seldom interact with the doll dragon in this way before.” (#2-6)

<table>
<thead>
<tr>
<th>Table 7. Ranking results of the interactivity aspect</th>
</tr>
</thead>
<tbody>
<tr>
<td># of students</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>0 (0%)</td>
</tr>
</tbody>
</table>

Discussion

Impact of tangible technology

The influences of the tangible version on student perceptions are summarized in Table 8. In short, students deemed the tangible agent system to be beneficial in terms of three aspects: attention, emotion, and interaction. It demonstrated that tangible agents have a positive influence on student learning.

<table>
<thead>
<tr>
<th>Table 8. Summary of the results in the tangible version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimensions</td>
</tr>
<tr>
<td>------------</td>
</tr>
<tr>
<td>Cognitive</td>
</tr>
<tr>
<td>Affective</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Behavioral</td>
</tr>
</tbody>
</table>

In a sense, the interface can be regarded as a channel that determines how much power or benefits students can gain from computer systems, including educational systems. Different from virtual user interfaces, tangible user interfaces might bring two possible potentials: psychological support and multisensory learning. Regarding psychological support, the experiences resulting from real touching on physical animal companions (e.g., using hands to touch their animal companions or physically playing ball with them) might attract students’ attention, and further enhance the level of pet ownership and emotional attachment (Melson, 2001). This might be the reason why the aspects of attention and emotion would be enhanced. Regarding multisensory learning, students deemed that the tangible system contributed to students’ attention, and is also beneficial to system interactivity. This result seems to echo the potential of physical and tangible technologies in learning: multiple interactions driven by visual stimulation and tangible sense can benefit student learning (Xu, Read, Sim & McManus, 2009). Different from virtual user interfaces, physical user interfaces might offer various ways (e.g., touch-based patting and interaction with animal companions) for students to interact with systems, creating a multisensory way to sustain students’ attention and behavioral reactions.

On the other hand, although the tangible system seemed to benefit students’ perception in terms of affective and behavioral aspects, such advantages did not further contribute to perceived helpfulness in the cognitive facet. Two possible reasons might explain this result. One is the limitation of use time. If this is the case, more long-term studies are required to investigate the consistency with the current study. The other is the lack of
incorporation of effective pedagogies, which is the optimal goal of animal companions extending the concept of learning companion systems (Chan, 1996) from virtual world to real world. A number of efficient pedagogies with educational agents have been demonstrated, such as one-to-one tutoring (Heift, 2015; Hernández et al., 2015), and collaborative learning (Ogan et al., 2014). Thus, based on the preliminary exploration of this study, more future work to integrate appropriate pedagogies is needed to enhance effectiveness of tangible technologies in educational systems.

From usability to enjoyment

From the perspective of human-computer interaction, usability is a significant consideration. According to Nielsen’s principles (Nielsen, 1993), usability has focused on aspects related to function, such as memorability, learnability, efficiency, error prevention, and satisfaction. Nevertheless, enjoyment is another factor that should be also taken into account. Some pioneering researchers have conducted related work on enjoyment. For instance, Malone (1980; 1982) proposed some heuristics for the design of enjoyable interfaces. Norman (2002; 2004) emphasizes the emotional design that could offer users enjoyable experience. Thus, a wider design perspective of “from usability to enjoyment” is advocated (Blythe et al., 2003). In addition to functionality, how enjoyable it is for students to interact with a learning system is also a critical consideration. In short, system usability and user enjoyment are two characteristics that should be concerned in human-computer interaction (Thomas & Macredie, 2002), but it is unclear to what extent an interactive educational system should incorporate enjoyment element based on a well-developed system usability.

The finding of this study revealed that the tangible animal companions could enhance students’ attention, and offer a highly interactive way to interact with animal companions. When students’ attentions and emotions are fully involved in the interaction, they are easier to go into the zone of deeper engagement (Csíjszentmihályi, 1990). In a sense, the strategy of animal companions used to enhance students’ interactivity (e.g., caring, learning, and assessment phases) is closely related with gamificated elements for creating their enjoyment experiences. Although the system usability might be thus decreased resulting from the complexity of tangible technology (e.g., involvement of various sensors and devices), their attention and emotion are thus stimulated. In other words, this study offers an experience in system development: animal companions with emerging technologies might serve as a lever to balance the usability and enjoyment, aiming to make student learning in both effective and enjoyable.

Conclusions

In response to the first research question (i.e., How can we integrate animal companions with tangible technology as a learning system with tangible representation style?), this study developed a My-Dragon system, where students interact with the system through a stuffed toy with tangible technology. In response to the second research question (i.e., What are student perceptions for different representations of animal companions among static, animated, and tangible styles?), the My-Dragon system is used as an example and its influence is further evaluated by comparing it with other non-tangible systems in terms of cognitive, affective, and behavioral dimensions. The results showed that such a tangible system is beneficial to students’ attention, emotion, and also contribute to perceived interaction. However, further investigation is required because of the limitations of this study. First, although some positive influences of tangible animal companions on the perception of elementary students were found, this is a pilot study with a small sample size. The effects with larger sample sizes should be further investigated. Second, In addition to the influence on students’ perception, the influence on learning effectiveness should be further investigated in the future.

Acknowledgments

The authors would like to thank the Ministry of Science and Technology of the Republic of China, Taiwan, for financial support (MOST-103-2511-S-155-002).

References


Guest Editorial: Trends and Research Issues of Learning Analytics and Educational Big Data

Gwo-Jen Hwang¹, Daniel Spikol² and Kam-Cheong Li³

¹National Taiwan University of Science and Technology, Taiwan // ²Malmo University, Sweden // ³Open University of Hong Kong, China // gjhwang.academic@gmail.com // daniel.spikol@mah.se // kcli@ouhk.edu.hk

Advancement of learning analytics and educational big data

Learning analytics and educational big data refers to the analysis and interpretation of educational data, such as the logs recorded in learning management systems, the interactive contents recorded in online discussion forums, or the learning process captured on video, to provide constructive feedback to learners, instructors or educational policy makers (Hwang, Hsu, Lai, & Hsueh, 2017; Hwang & Wang, 2004; Siemens & Long, 2011). These technologies, processes, and analyses play an important role of optimizing students’ outcomes, teachers’ learning designs as well as the improving the learning environment (Greller & Drachsler, 2012; Hwang, Chu, & Yin, 2017).

In the past decade, the number of applications and studies related to learning analytics and educational big data has been increased at a fast pace. By searching for the publications in the Scopus data from 2008 to 2017, it is found that the number of publications that match the keywords “learning analytics” or “educational big data” is increased from 0 in 2008 to 534 in 2017, as shown in Figure 1. This implies that learning analytics and educational big data has become a vital research trend in educational technology.

Figure 1. Number of publications related to learning analytics and educational big data from 2008 to 2017

Figure 2. Major contributing countries for learning analytics and educational big data studies
By categorizing those publications by the authors’ nationality, the number one contributing country is the United States, followed by Spain, United Kingdom, Australia, Germany, Canada, and Netherlands, as shown in Figure 2. However, from the statistical data, we see that researchers from many countries have contributed to the relevant studies.

By categorizing the publications by the applied subjects, it is found that the learning analytics and educational big data approach has been applied to the analysis of the data collected from a variety of courses, as shown in Figure 3. The most frequently applied courses are computer science, followed by social sciences, engineers, and mathematics, while few applications were related to Agricultural and Biological Sciences, Microbiology, Multidisciplinary, Nursing, Chemical Engineering, Dentistry, and Neuroscience, showing the growing space of applying the approach in the future.

![Figure 3. Subjects of the studies related to learning analytics and educational big data](image)

**Publications in this special issue**

In this special issue, 14 quality papers were accepted for publication. Most studies analyzed students’ learning behavioral patterns for providing suggestions to teachers or researchers for developing better learning strategies or systems. For example, in the study conducted by Hwang, Chen, Chen, Lin, and Chen, the students’ self-reflective learning instances and peer-reflective learning instances were analyzed to compare the effectiveness of two learning models; in the study conducted by Mouri, Uosaki, and Ogata, a learning analytics approach was employed for supporting ubiquitous language learning via e-books based on the analysis on students’ learning logs. Another study conducted by Chen, Liu, and Shou analyzed the students’ behavioral patterns to explain why those learning in the non-competitive gaming environment had better learning performances than those learning in a competitive one. Sun, Hwang, Lin, Yu, Pan, and Chen analyzed the behavioral patterns and brainwave data of the students learning with a concept mapping approach for providing suggestions to teachers who intend to employ the approach based on voting. Yang and Chu analyzed students’ behavioral patterns in a progressive prompting-based gaming environment for mathematics.

In addition, several studies have attempted to analyze the behaviors of students in collaborative, peer-assessment or peer-tutoring activities. For example, Zhang, Zou, Huang, and Zhang analyzed group behavioral data collected from a long-term and large-scale online course; Ma, Xin, and Du conducted a peer coaching-based professional development program and analyzed the learning participation of those participated teachers; Hsu and Chang analyzed students’ behavioral patterns in a peer assessment-based webpage design activity. Lin and Hwang investigated the factors affecting EFL students’ oral performance in a flipped classroom by analyzing students’ interactive content on Facebook.

Learning analytics can also be used to provide personalized supports by analyze and predict students’ learning problems and needs. For example, Lu, Huang, Huang, Lin, Ogata, and Yang applied a learning analytics approach
to derive early prediction for students’ academic performance in a blended learning context; Zou and Xie tried to analyze students’ learning status for providing personalized guidance in English vocabulary learning; Choi, Lam, Li and Wong employed a learning analytics approach to predict the potential at-risk students.

On the other hand, the data from open learning environments can provide very useful information for learning design and policy making. For example, Wu, Yu and Wang tried to analyze students’ learner interests in open learning environments using a learner-topic model; Bey, Jermann and Dillenbourg analyzed the data collected from MOOCs for comparing the performances of two assessment models.

Potential research issues

The aim of this special issue is to enable researchers to see the up-to-date development of learning analytics and educational big data studies. The papers published in this special issue provide a good reference for those who intend to conduct educational technology research from a new perspective. To provide further, suggestions to researchers, the guest editors of this special issue summarize the potential research issues of learning analytics and educational big data in the followings.

- Modifying the existing methodologies or proposing new ones to improve the efficiency of learning analytics and educational big data analysis
- Employing learning analytics and educational big data approaches in various application domains; in particular, those seldom investigated ones.
- Using learning analytics and educational big data approaches to investigate the factors affecting students’ learning performances
- Providing personalized supports by analyzing students’ learning logs and making predictions
- Comparing the behavioral patterns of the students with different achievement levels and providing suggestions to those low-achievement ones.
- Investigating the impacts of different learning strategies or systems on students’ behavioral patterns.
- Investigating students’ learning behaviors when learning with new technologies.
- Investigating the correlations between students’ behaviors, learning perceptions and performances.

References


Learning Behavior Analysis of a Ubiquitous Situated Reflective Learning System with Application to Life Science and Technology Teaching

Wu-Yuin Hwang¹, Hong-Ren Chen², Nian-Shing Chen³, Li-Kai Lin¹ and Jin-Wen Chen²

¹Graduate School of Network Learning Technology, National Central University, Taoyuan, Taiwan // ²Department of Digital Content and Technology, National Taichung University of Education, Taichung, Taiwan // ³Department of Information Management, National Sun Yat-Sen University, Taiwan // wyhwang@cc.ncu.edu.tw // hrchen@mail.ntcu.edu.tw // nschen@mis.nsysu.edu.tw // //y2j025qq@cl.ncu.edu.tw // insistene99@gmail.com

Abstract

Education research has shown that reflective study can efficiently enhance learning, and the acquisition of knowledge and skills from real-life situations has become a focus of interest for scholars. The knowledge-learning model based on verbal instruction, used in traditional classrooms, does not make use of real-life situations that encourage students to engage in reflective learning. However, by implementing the Ubiquitous Situated Reflective Learning System (USRLS), learners can be provided with real situations, faced in daily life at any time, to encourage them to engage in reflective learning with regard to information pertinent to the class. This study adopted a quasi-experimental design to assess the efficacy of the two learning models. The research subjects were 52 students from two grade 5 classes in one elementary school in the middle part of Taiwan. The USRLS was used for teaching the experimental group, while the traditional oral teaching method was used for the control group. The learning content of fifth-grade life-science technology classes consists of units on burning and rusting in the context of the life sciences and technology. The research results showed that (1) the learning effectiveness of the USRLS is superior to that of the traditional oral teaching model and (2) students in the high-learning achievement (HLA) group are best suited to a text-based self-reflective learning strategy, while students in the low-learning achievement (LLA) group can obtain more help by using a text-based peer-reflective strategy. Students noted that the learning cycle of a situated reflective learning model encouraged them to consider lesson content, help them review their answers, and enabled them to increase their focus on the concepts and information in the learning task.

Keywords

Situated learning, Reflective thinking, Ubiquitous application

Introduction

Situated learning is a theory of learning that stresses the role of context. According to this approach, learning includes the situational context in which it occurs (Brown, Collins, & Duguid, 1989; Hou, 2011). Indeed, knowledge is embedded in its situational context as well as in the learning activities (Kim & Hannafin, 2011). Situated learning emphasizes the non-official or incidental learning that occurs outside the classroom; it views actions during real interactions as ways to acquire knowledge and capabilities (Zurita, Baloian, & Frez, 2014). Situated learning includes not only the cognitive process of knowledge acquisition but also the learning that occurs during social interactions (Saigul, 2012; Hwang, Chen, Shadiev, Huang, & Chen, 2012). It uses real environments to encourage independent and autonomous thinking as well as to apply active methods to acquire knowledge, thus emphasizing that learning should be based on real-life situations (Arnseth, 2008). Based on learning situations in which different academic subjects are taught, situated learning can include apprenticeships, collaborative learning, multiple opportunities for practice, and the articulation of what has been learned, all of which have been proven to enhance learning achievement (Compton, 2013; Woolf & Quinn, 2009).

Reflection on learning can train students to gather information from their everyday lives and to apply it to solving problems by using their own knowledge to solve life technology-based problems and by relying on scientific evidence to explain the outcomes (Greiff, Holt, & Funke, 2013). Reflection has an important role in the learning process (Hung, Yang, Fang, Hwang, & Chen, 2014; Koong, Yang, Wu, Li, & Tseng, 2014). When learners are able to reflect on the instructional materials provided during the learning process, they are able to gain a better understanding of their effects (Aleven, McLaren, Sewall, & Koedinger, 2009; Chen, Kinshuk, Wei, & Liu, 2011). Studies have shown that learners who encounter real situations while reflecting on their learning show improved learning effects (Kim, 2011; Russell, 2008).

During the implementation stage of the sciences and technology curriculum guidelines, fifth-grade students cultivate their abilities to observe and classify knowledge (Hsu & Kuan, 2013). Verbal instruction employed in traditional classrooms does not provide real-life situations for students to experience and explore. Facilitating
elementary school students to engage in thinking activities that involve reflective learning can improve their observation and classification abilities through applying their classroom knowledge. This study proposes a learning strategy based on a situated reflective learning model and system. This system provides learners with time to reflect on and to share the knowledge they gain in their daily lives in response to real-life situations. This study focused on the science and technology curriculum of fifth-grade students; the instructional design was based on the learning cycle of a situated, reflective learning model (Collins, 1994; Zimmerman & Schunk, 1989) to elucidate its educational benefits and to analyze differences in the learning behaviors of high-achieving and low-achieving students. Therefore, the research questions of this study are as follow: (1) The study analyses the learning effectiveness differences for the USRL teaching and traditional oral teaching. (2) The present study explored learning behavioral differences between the HLA group and the LLA group in self-reflective learning and in peer-reflective learning. (3) The study also investigated the correlation between learning achievements and reflective learning behaviors of the HLA group and the LLA group.

Literature review

Situated learning

Many positive research outcomes have been associated with the introduction of situated learning theory, which has been most widely applied in the natural sciences. Onsite observations for purposes of annotation and comparison can help students understand the scientific phenomena described in abstract text (Chu, Hwang, & Tsai, 2010; Tan, Lin, Shu, & Liu, 2012). In learning Chinese, situated learning theory can be used to combine Chinese poetry that is difficult to understand with situations that deepen understanding of the meanings of the poetry (Chen & Lin, 2016; Shih, Tseng, Yang, Lin, & Liang, 2012). In learning mathematics, Shih, Kuo, and Liu (2012) used a learning method involving a mathematics path to connect related mathematical concepts with objects in daily life to help learners understand abstract concepts. These studies have consistently shown that using real situations to elucidate knowledge that is otherwise difficult to understand can achieve positive learning effects. Hwang, Shi, and Chu (2011) believed that if there is no appropriate learning strategy or tool to help learners with situated learning, the learning effects are usually disappointing, despite the use of novel learning strategies that combine new situations with E-learning.

The instructional strategies used in situated learning approaches help students learn by observing and participating in real situations and emphasize that learning involves active interactions between learners and the environment that integrate previous experiences in the service of learning new knowledge or developing problem-solving strategies. Collins (1994) proposed that situated learning strategies involve six characteristics as follows: (1) Authentic context: all knowledge and skills must be learned in real situations to help students immediately apply the knowledge they learned to their lives. (2) Coaching: intersperse learning between the completion of learning tasks to develop familiarity with specific knowledge or skills so that students learn to solve specific problems at the same time as they learn to apply knowledge in different situations. (3) Articulation: guide students to engage in deeper thinking about knowledge they have learned so they can generalize it to other situations. (4) Reflection: guide learners to engage in reflecting on whether what they have learned is accurate and to think about whether there are other solutions. (5) Circulation: provide consistent opportunities for learning so that learners can continuously practice learning similar material and improve their problem-solving abilities, leading to a sense of accomplishment with regard to the process of solving problems. (6) Multiple media: use different types of media to demonstrate different learning characteristics so that the context of instruction more closely resembles actual contexts, which enhances the effects of learning.

Reflective learning

Reflective ability has been seen as a learning strategy and means to elevate learning benefits (Hsu, 2011). Reflection can be used to evaluate learning accomplishment experiences and promote the elevation of learning motivation (Mansvelder-Longayroux, Beijaard, & Verloop, 2007). Barrett (2008) pointed out that learning systems with the e-portfolios function, and design and completeness of reflection mechanisms would affect the reflective techniques and thinking of learners. Recently, many studies on the reflective learning issue have explored the effect of reflection on professional knowledge learning in university students, finding that reflection is positive for learning, and promoting and encouraging students to undertake reflective learning (Bhattacharya & Chauhan, 2010; Koong et al., 2014; Van de Boom, paas, & Van Merriënboer, 2007).
Common literature on specific exploration of reflective learning structures includes self-reflection in the self-regulated learning theory by Zimmerman and Schunk (1989) and continuum of levels of reflection proposed by Grossman (2008). In the self-regulated learning theory, self-reflection can be divided into four cycling stages. In the self-evaluation stage, learners focus on their own learning performance to compare them with those of others, and evaluate the reflection of their mistakes. In the attributions stage, learners reflect on their mistakes and think about the reasons. In the self-reaction stage, learners derive different self-reflection due to different attribution conclusions, if learners acquire positive self-reactions, learners would consider feasible solutions, and effectively improve upon the incorrect concepts. In the adaptive stage, the attributions stage affects the outcomes of the self-reaction stage. It promotes learners’ discovery and confirmation of learning mistakes, so that they take action to adjust the learning actions of learners.

Previous literature related to situated and reflective learning have discussed differences in learning effectiveness for students with higher and lower achievement. For instance, Lin (2014) showed that the progress of students with lower achievement in geometry was superior to those with higher achievement, with the assistance of peers, which is of particular help in facilitating our knowledge of learning achievement based on measurement and estimating ability. According to Chen and Lin (2016), their situated learning game system in Chinese poetry focused on poetry of the Tang Dynasty, where the difficulties encountered by poets during the writing process are simulated into animation. Moreover, the multimedia annotation e-book learning system, to facilitate English language learning, can provide personalized learning, whereby annotations are made regarding more complex concepts to assist in follow-up study at home. Additionally, there is an annotation-sharing feature so that low and high achievement students can see each other’s comments, which contributes to the learning progress. It has been documented that high achievement students engage in annotation sharing significantly more than their low achievement counterparts (Hwang, Liu, Chen, Huang, & Li, 2015).

**Ubiquitous situated reflective learning**

**Situated reflective learning model**

The situated reflective learning model was divided into two phases: the situated reaction and the reflective reaction. The situated reaction was based on the situated instructional strategy recommendations offered by Collins (1994), whereas the reflective reaction was based on the integration of the self-reflective steps outlined in the self-regulated learning theory developed by Zimmerman and Schunk (1989), as shown in Figure 1. This model was divided into five steps. The first step is articulation, in which classroom teachers lead students in making judgments and classifications, enabling them to think more deeply, thereby facilitating knowledge transfer. The second step is authentication, which involves the use of articulation to derive knowledge and skills from real-life environments; in this step, students learn to observe and record their thinking processes as they acquire knowledge. The third step is evaluation, which focuses on the knowledge acquired in the authentication step via reflection on the accuracy of the knowledge gained; it enables students to evaluate the correctness of their thinking or responses. Students discuss and share ideas with classmates to facilitate the process of evaluating knowledge acquisition. The fourth step is planning, which uses the evaluation and reflection from the previous steps to consider whether knowledge had been misclassified or if the concepts are incorrect; students are encouraged to consider which real-life scenarios match these concepts and to re-evaluate their plans. The fifth step is adaptation, which confirms the reasons for erroneous learning and leads to adjustments in learners’ behavior. In other words, the plan from the previous step is converted into real action, which, in turn, increases students’ awareness of the knowledge they have acquired.

The implementation steps of the situated reflective learning model were based on the teaching process of Taiwan’s elementary schools. One reflective cycle process was conducted according to the connection generated by elementary school students between life situation topics and knowledge learned in the classroom. This model adopted the advantages of the situated learning theory proposed by Collin (1994) and the self-regulated learning theory proposed by Zimmerman and Schunk (1989) and simplified them into five steps. Situated learning theory emphasizes the link between the real-life environment and knowledge learned in class, as well as the ability development of its immediate use, but there are no detailed steps or implementation models in the narrative of reflective learning, so the steps of self-regulated learning theory were added to establish a complete situated reflective learning model. In self-regulated learning theory, it is thought that the reflective learning process requires specific practice to make the learning adjustment. Many scholars believe that a reflective learning process requires definitive ways to adjust learning (Dabbagh & Kitsantas, 2012; Koong et al., 2014; Littlejohn, Milligan & Margaryan, 2012; Montgomery, 1993). Therefore, step four of the five steps was a plan to evaluate
classification errors or concept errors and redraft a plan to carry out correct learning. Such a model can be fully integrated with the teaching of elementary school courses.

Figure 1. The five steps of situated reflective learning model

Design of ubiquitous situated reflective learning system

The ubiquitous situated reflective learning system was developed based on the above proposed situated reflective learning model, which allows for reflective learning about life context and classroom knowledge at any time and place through mobile devices. This system includes an E-book for the course, teacher lectures, annotations, reflective learning, and instructional activities. For the function of E-book for the course, this system consists of an E-book for the units pertaining to fire and rust in the life technology class as shown in Figure 2.

Figure 2. Ebook for the course

For the function of teacher lecture, students can use the lecture function to record the lecture for review at home as shown in Figure 3. For the learning annotation, students can annotate their class notes in the E-book in the learning process. This includes not only text annotations but also vocal comments and photographs. Annotations can be saved or deleted as shown in Figure 4. For the reflective learning, students can use the system to browse the learning content drawn from the life contexts of other classmates and engage in online discussions about and validation of answers, as shown in Figure 5. For learning activities, these are matched to the curricular instruction so that students can engage in reflective learning through connections with their life contexts.

In the step of establishing a new object, the system uses the GPS positioning function to search for the current location of the student to record data for the experiment, as shown in Figure 6. In the step of taking photographs, after students input and identify the names of objects, they select photographs of them, and the system again confirms whether the object name is correct. If they choose “confirm,” the camera function starts, and students point to the center of the screen at the new object to take a photograph of the object. If the photograph is blurry or distorted, they can retake it, as shown in Figure 7. In the step of description and recording, students can choose whether to input text or a recorded verbal description regarding why the photo conforms to the knowledge learned in the course. When they finish the explanation, they press “complete” to send it, as shown in Figure 8.
In terms of USRLS instructional activity functions, the steps in the situated reflective learning model assist reflective learning in relation to real-life situations, as shown in Table 1.

<table>
<thead>
<tr>
<th>Step</th>
<th>Purpose description</th>
<th>USRLS supporting functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Articulation step</td>
<td>Learners thinks about learning, then judges and classifies</td>
<td>Teachers with explanatory function</td>
</tr>
<tr>
<td>Authentic step</td>
<td>Discover connections in knowledge from real-life situations</td>
<td>Learning annotation functions, gps situated trigger function, photography function, and sound recording function</td>
</tr>
<tr>
<td>Evaluation step</td>
<td>Learners can reflect on the correctness of knowledge and discuss with other students</td>
<td>Learning annotation function, reflective learning function, photography function, and sound recording function</td>
</tr>
<tr>
<td>Plan step</td>
<td>Reflection processes to confirm errors in concepts, and reestablish the plan to implement learning of correct knowledge.</td>
<td>Learning annotation function and reflective learning function</td>
</tr>
<tr>
<td>Adaptation step</td>
<td>To confirm the reason for incorrect learning, to covert plans to actual action</td>
<td>Learning annotation function and reflective learning function</td>
</tr>
</tbody>
</table>

**Method**

**Participants and learning material**

The sample consisted of 52 students in two fifth-grade classes at an elementary school in central Taiwan. One class of 25 students formed the experimental group and another class of 27 students formed the control group. Students had access to information technology tools, including word processing, presentations, and tablet computers. The instructional material consisted of units on the topics of fire and rust, which were part of the life technology course. In the fire unit, students learned about the classification of fires, such as fires caused by flammable solids, liquids or gases, and electrical equipment. In the rust unit, students learned about different types of rust, including the rusting properties of iron, aluminum, and copper. The fire and iron are selected as the learning unit because in reality they are commonly seen even to the elementary school students in daily life. It can be used to conduct the situated reflective learning activity. The elementary school students in experimental group and control group come from the same elementary school, have the same background knowledge and also share the same teacher in the course of life science and technology.

**Research design**

This study adopted a quasi-experimental design. The control group received traditional oral instruction, whereas the experimental group received instruction using the ubiquitously situated, reflective learning system (USRLS). Figure 9 shows the research framework diagram. The independent variable was that of the different instructional methods. The experiment group used the USRLS instructional method, while the control group received traditional oral instruction. The control variables included instruction by the same teacher, the same instructional times, the same instructional progress, the same content, and the same test methods. The covariate variable was the learning achievement pre-test score. The moderator variable was the level of learning achievement. HLA group and LLA group are distinguished by students’ academic performance in the previous semester. The top 50% students are selected as HLA group while the last 50% students are selected as LLA group. Such analysis method with different learning ability has been adopted by many literatures (Chen & Lin, 2016; Lin, 2014).

The dependent variables were learning achievement post-test scores, the number of self-reflective learning instances, and the number of peer-reflective learning instances. To ensure the reliability of the experimental outcomes, the two groups were both taught by the same teacher during the 4-week instruction period, and both groups followed the same instructional schedule. The first 2 weeks of the life technology course covered the fire unit, and this was followed by the rust unit in the second 2 weeks. The control group received traditional oral instruction from the teacher to enable them to engage in reflection regarding the curricular content, whereas the experimental group used the USRLS to participate in context-reflective instructional activities based on the same content. The five steps in the situated reflective learning activity belong to the situated reflective learning model proposed, so that students can reflect on the knowledge seen in the daily life situations and learned in the classroom and transform the incorrect concept into the correct knowledge. This activity cooperates with the
teaching progress in elementary school and a total of 4 weeks of learning time is planned in the theme learning of fire and iron. For example, students were asked to connect what they had learned with what they had experienced in life, to reflect on the causes of fires, and to record and photograph them; they were also encouraged to discuss their ideas with their classmates to confirm their answers. One question, for example, concerned the type of fire frequently caused by faulty electrical equipment. This learning system encouraged students to discuss, evaluate, and reflect on the content with their classmates to determine whether they had made erroneous classifications of, and thought processes about, electrical equipment. In the unit about rust, students were asked to connect with, and reflect on, experiences in their daily lives to help them to classify the rust found on metal products they encounter regularly as shown in Figure 10.

The control group received traditional oral instruction. The situated reflective learning model was employed for the experimental group, and records on paper were used to implement the five steps (listed below). Iron rust was used simply for explanation.

Step 1: Able to check the textbook and use knowledge taught by the teacher to evaluate and classify knowledge on iron rust.

Step 2: Able to connect learned knowledge with life contexts, identify which objects will rust, and write down explanations for the objects.

Step 3: Able to think about the casing of fire alarms commonly seen in daily life and what materials they are made from that may rust. When students next return to the classroom, they take out their reflective learning sheet to help them carry out discussion with students in their group, and they evaluate and reflect on whether their classifications and ideas about fire alarm casings are correct.
Step 4: If they are wrong, students set out to reconsider which other objects also rust, look for them in their daily life contexts, and record them on their reflective learning sheets.

Step 5: Able to review the textbook and knowledge taught by the teacher on iron rust to confirm the reason for any incorrect learning and thereby improve the accuracy of their knowledge.

Instrument of assessing students’ reflective level

Learners who engage in reflection, based on the knowledge they have learned, show better memory and understanding. Based on the reflective content of learners, Bain (1999) proposed the following five-level criteria for the evaluation of reflection: reporting, responding, relating, reasoning, and reconstructing. Reporting refers to merely repeating existing content. Responding refers to using only a few concepts or explaining only phenomenological facts, rather than causes, and describing only personal thoughts. Relating refers to describing the relationships in the text and explaining the reasons behind them. Reasoning refers to an in-depth explanation of reasons for a phenomenon and discussion of the relationships between theory and practice. Reconstructing refers to the ability to describe a high-level reasoning and reconstruction process in the context of both personal experiences and more general rules in a way that demonstrates a systematic understanding of the theory and a process by which conclusions are reached. Chen, Wei, Wu, and Uden (2009) believed that some learners’ reflective content is too simple, incomplete, or incorrect with regard to an understanding of a pertinent problem. However, their original schema did not include an appropriately low level into which such content could be classified. Therefore, the following three reflective levels were added: “nonsense,” “too simple,” and “incomplete description.” This study used the evaluation criteria for reflection proposed by Chen, Wei, Wu, and Uden (2009) to analyze the level of reflection demonstrated by learners regarding the life science technology course.

Results

Analysis of learning achievement

This section addresses the influence of different teaching methods on learning achievements and aims to assess differences in achievement associated with the USRLS teaching approach versus traditional lecturing. The results of a one-way analysis of covariance (ANCOVA) with regard to within-group homogeneity are presented in Table 2. The F value is 0.171, and the p-value is .681. It did not reach the level of significance. The slopes of the regression lines for the two groups can be regarded as the same. Table 3 shows that after the effect of a covariate was eliminated, the F-value was 87.512 and the p-value was .000, which reflect statistical significance at the level of .05 and indicate a significant difference between the learning outcomes associated with the USRLS teaching approach and traditional lecturing. Table 4 shows that, after adjustment, the average score of the experimental group on the learning achievement test was 82.462, which was higher than the adjusted average of control group, 76.017. Thus, the effectiveness of the USRLS approach to teaching fifth-grade science and technology classes was superior to that of traditional lecturing.

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of squares</th>
<th>df</th>
<th>Mean square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within-group × covariate</td>
<td>7.203</td>
<td>1</td>
<td>7.203</td>
<td>0.171</td>
<td>.681</td>
</tr>
<tr>
<td>Error</td>
<td>2016.520</td>
<td>48</td>
<td>42.011</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of squares</th>
<th>df</th>
<th>Mean square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contrast</td>
<td>3614.289</td>
<td>1</td>
<td>3614.289</td>
<td>87.512</td>
<td>.000</td>
</tr>
<tr>
<td>Error</td>
<td>2023.723</td>
<td>49</td>
<td>41.300</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>Pretest mean</th>
<th>Posttest mean</th>
<th>Adjusted mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment Group</td>
<td>83.24</td>
<td>82.56</td>
<td>82.462a</td>
</tr>
<tr>
<td>Control Group</td>
<td>83.04</td>
<td>75.93</td>
<td>76.017a</td>
</tr>
</tbody>
</table>

Note. * = Covariates appearing in the model are evaluated at the following values: pretest = 83.13.
Analysis of the self-reflective learning behavior of HLA and LLA students

This section addresses differences between the quality of text-based self-reflection (QTSR) and that of quality of voice-based self-reflection (QVSR) in high-level-achievement (HLA) and low-level-achievement (LLA) students who were taught using the USRLS teaching approach. According to the results of the independent-sample t-tests, presented in Table 5, HLA students scored significantly higher than did LLA students ($p < .001$) with regard to the QTSR. This result shows that HLA students used the situated reflective model provided by the system more effectively than did the LLA learners with regard to improving their knowledge and providing correct answers. However, we found no significant difference between HLA and LLA students with regard to the QVSR ($p = .471$). This may be due to class rules and teacher preferences that favor students’ vocalizing of self-reflections. It may also be due to the fact that there is too much noise in real-life situations, which renders learners unable to use the VSR in real time. Figure 11 shows that the fact is not that the QTSR of HLA learners has gone up after using USRLS teaching, instead, it is that the QTSR of LLA learners has declined gradually as time passes by. This might be due to the fact that LLA learners’ feeling of freshness about USRLS system in the beginning of the experiment. However, as this feeling of freshness fades away, low achievement learners’ QTSR declines as well.

Table 5. Independent-sample t-test analysis for self-reflective learning behavior

<table>
<thead>
<tr>
<th>Learning achievement</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quantity of text-based self-reflection</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>12</td>
<td>5.24</td>
<td>2.57</td>
<td>0.67</td>
<td>.934</td>
</tr>
<tr>
<td>Low</td>
<td>12</td>
<td>2.74</td>
<td>0.62</td>
<td>0.73</td>
<td>.471</td>
</tr>
<tr>
<td><strong>Quantity of voice-based peer-reflection</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>12</td>
<td>5.57</td>
<td>2.98</td>
<td>0.74</td>
<td>.473</td>
</tr>
<tr>
<td>Low</td>
<td>12</td>
<td>4.37</td>
<td>0.47</td>
<td>1.14</td>
<td>.001</td>
</tr>
</tbody>
</table>

Note. *p < .05; **p < .01; ***p < .001.

Figure 11. QTSR for HLA and LLA

Analysis of the peer-reflective learning behavior of HLA and LLA Students

This section discusses differences in the quality of text-based peer-reflection (QTPR) and the quality of voice-based peer-reflection (QVPR) in HLA and LLA students who were exposed to the USRLS teaching approach. We will focus on how learners used the situated reflective learning mode provided by the USRLS system to assess the content presented by other students to improve their own understanding of the material. Table 6 presents the results of the independent-samples t-test, which show that HLA learners performed better than LLA learners in terms of both QTPR ($p < .001$) and QVPR ($p < .001$). This result indicates that HLA students made better use of the situated reflective learning mode provided by the system. Indeed, Figure 12 shows that the QTRP of LLA students was not as high as that of HLA students, but that the QTPR of LLA students gradually increased as a function of time. This might be because LLA students observed the reflected content of other students (especially that of HLA students) and were able to improve their understanding and make better use of the USRLS system.

Table 6. Independent-sample t-test analysis for peer-reflective learning behavior

<table>
<thead>
<tr>
<th>Learning achievement</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quantity of text-based peer-reflection</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>12</td>
<td>5.57</td>
<td>0.74</td>
<td>6.473</td>
<td>.000</td>
</tr>
<tr>
<td>Low</td>
<td>12</td>
<td>4.37</td>
<td>0.47</td>
<td>4.314</td>
<td>.001</td>
</tr>
<tr>
<td><strong>Quantity of voice-based peer-reflection</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>12</td>
<td>5.57</td>
<td>0.74</td>
<td>6.473</td>
<td>.000</td>
</tr>
<tr>
<td>Low</td>
<td>12</td>
<td>4.37</td>
<td>0.47</td>
<td>4.314</td>
<td>.001</td>
</tr>
</tbody>
</table>

Note. *p < .05; **p < .01; ***p < .001.
Correlation between reflective learning behavior and learning achievement

This section discusses the connection between learning achievement and reflective learning behaviors of HLA and LLA learners. Reflective learning behaviors include QTSR, QVSR, QTPR, and QVPR. Table 7 shows that HLA learners are highly correlated to QTSR, while LLA learners are highly correlated to QTPR. This means that good QTSR can be transformed effectively into improvements in learning achievement for HLA students but that QTPR is a crucial factor for knowledge learning among LLA students.

Table 7. Pearson correlation between learning achievement and reflective learning behavior

<table>
<thead>
<tr>
<th>Learning achievement</th>
<th>QTSR</th>
<th>QVSR</th>
<th>QTPR</th>
<th>QVPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA</td>
<td>.762*</td>
<td>.267</td>
<td>.426**</td>
<td>.348*</td>
</tr>
<tr>
<td>LLA</td>
<td>.323*</td>
<td>.070</td>
<td>.757**</td>
<td>.117</td>
</tr>
</tbody>
</table>

*Note.* *p < .05; **p < .01; ***p < .001.

Discussion

The results of Tables 3 and 4 showed that the experimental group, which received the URSLS teaching intervention, performed significantly better than the control group, which received traditional oral teaching. This result confirms previous research suggesting that reflective learning can improve learning. Hung et al. (2014) argued that context-aware reflective learning strategies based on videos can improve learners’ levels of reflection and provide real-time individual guidance in the application of reflective learning. Based on evaluations undertaken with vocational education software, Koong et al. (2014) believed that different reflective learning strategies can affect the acquisition of knowledge-based skills.

These data of Tables 5 and 6 show that HLA learners can effectively use the situated reflective learning mode provided by the system to improve their learning achievements. Previous studies on reflective learning did not address the effect of the use of real-life situations to guide students’ reflective learning, self-reflective learning behaviors, and peer-reflective learning behaviors on learning achievements. This study found that HLA students performed better using self-reflective learning strategies. With respect to peer-reflective learning behaviors, HLA learners showed greater improvement in their QTPR and QVPR compared with LLA learners. LLA learners spent more time using the USRLS system over time, and their QTPR improved. Additionally, the QTPR of LLA learners was strongly correlated with their learning achievements. It can, therefore, be concluded that LLA students found text-based peer reflective learning strategies particularly helpful. That is to say, students in HLA can understand fact phenomenon, connect personal experience, illustrate and verify the judgment reason for knowledge learned in class. Relatively, students in LLA can only reach incomplete level and reporting level, and they can only repeat the contents without extra ideas, and reflection contents are quite incomplete. The average score of reflective learning behavior in LLA Group is relatively lower from Table 5, showing that the students’ reflective learning in LLA Group will not enter responding stage or relating state easily. Under the peer-reflective learning strategies, students’ highest reflection level in HLA can reach reasoning level, and they can make a detailed explanation for the reasons why something happens; while students’ highest reflection level in LLA can reach responding level, and they can make a brief description for the reasons why something happens.
Conclusions

This study proposes a situated reflective learning model for the implementation of a reflective learning system. This model was applied in a fifth-grade science and technology class to enable students to think about and hone their knowledge they gained based on real-life situations in their daily lives. USRLS can quickly connect classroom knowledge to seen real-life situations for reflection and confirmation, and students can discuss observations in the situation. This situation triggered reflection is an important innovation in the present study. Our data showed that the experimental group, which received the USRLS teaching intervention, performed significantly better than the control group, which received traditional teaching. This study found that the self-reflective learning behaviors varied greatly between HLA students and LLA students using the USRLS. HLA learners performed significantly better than LLA learners with respect to self-reflective learning behaviors. Furthermore, the self-reflective learning behaviors of HLA students were strongly correlated with their learning achievements, whereas their peer-reflective learning behaviors were more weakly correlated with such achievements.

Aimed at the innovative contribution and effect of this paper, the supplementary description is conducted respectively in teaching practical level and academic innovative level. In terms of academic contribution, this paper proposed the situated reflective learning model, which is applied to the ubiquitous real-life situation, so as to strengthen the learners’ situated, reflective and self-regulated cyclic learning in face of the real-life situation. Meanwhile, HLA Group and LLA Group were used to analyze the learning behavior in this model, which can be used as the basis and application of follow-up researches related to situated reflective learning. From the perspective of instructional practice, the USRLS can help students use real-world experiences to validate their observations and organization of what they have learned in the classroom and to rethink how their answers differ from or resemble the narratives of other students. The USRLS enables the teacher to understand clearly the needs of each student through the annotations and information shared in reflective learning and to provide further group or individual instruction. Successful development of the USRLS system, which follows the recommendations of Hwang, Shi, and Chu (2011), will be very useful in implementing the ongoing educational reforms set forth in the grades 1–9 curriculum issued by the Taiwan Ministry of Education. This learning style, which involves the use of actual contexts and digital modalities, requires appropriate pedagogical strategies and usually leads to significant improvements in learning performance.

The study implemented the USRLS to provide the real-life situation reflection is currently merely used in the science learning in elementary school, and math can be designed for E-learning materials in the future. Thus, students can learn mathematical concepts from the experience in daily life to cultivate their problem-solving ability in math. In consideration of the current practical teaching status in elementary school in Taiwan, the real-life situations that students can discover seem to be limited, causing the issues shared between students to be overly concentrated. In the future of advanced plans, the USRLS will apply the technology of Augmented Reality to make students understand the thinking in other different situations, so that the discussion between students becomes deeper and more focused. In particular, students can improve their stimulation in learning knowledge and learning motivation, and students’ academic achievements can become apparently better in the strategy of cooperative learning.

References


Learning Analytics for Supporting Seamless Language Learning using E-book with Ubiquitous Learning System

Kousuke Mouri1, Noriko Uosaki2 and Hiroaki Ogata3

1Institute of Engineering, Tokyo University of Agriculture and Technology, Japan // 2Center for International Education and Exchange, Osaka University, Japan // 3Kyoto University, Japan // mourikousuke@go.tuat.ac.jp // n.uosaki@gmail.com // hiroaki.ogata@gmail.com

*Corresponding author

ABSTRACT

Seamless learning has been recognized as an effective learning approach across various dimensions including formal and informal learning contexts, individual and social learning, and physical world and cyberspace. With the emergence of seamless learning, majority of the current research focus on realizing a seamless learning environment at school or university. However, the utilization of the collected learning logs still remains a challenge yet to be explored. In this study, an e-book with ubiquitous learning system called SCROLL is developed to collect and analyze learning logs in the seamless learning environment. Moreover, this paper presents our analytics in contribution to bridging the learning between eBook learning and real-life learning. An experiment was conducted to evaluate (1) whether VASCORLL 2.0 (Visualization and Analysis System for Connecting Relationships of Learning Logs) is effective in connecting the words learned through eBook to those learned from real-life, and (2) which social network centrality is the most effective to enhance learning in the seamless learning environment. Twenty international students participated in the evaluation experiment, and they were able to increase their learning opportunities by using VASCORLL 2.0. Furthermore, the betweenness centrality was found useful in finding central words that bridge eBook and real-life learning.

Keywords
Mobile learning, Ubiquitous Learning, Seamless learning, Learning analytics, Ubiquitous learning analytics

Introduction

Recently, several researchers in language learning have focused on supporting in-class language learning and out-of-class language learning. In discussing language learning issues, it is often argued how recent technologies such as mobile and ubiquitous support language learning because of the lack of autonomous learning and authentic social interaction. Using mobile and ubiquitous technologies, learners can actively save what they have learned as learning logs anytime and anywhere, thereby sharing and making learning more collaborative.

With the emergence of these technologies, seamless learning is defined as an approach “when a person experiences a continuity of learning, and consciously bridges the multifaceted learning efforts across a combination of locations, times, technologies or social settings” (Wong et al., 2015, p. 9). Several researchers in the seamless language learning field have pointed out that mobile and ubiquitous technologies have enabled students to learn continuously across different contexts (Looi et al., 2015; Milrad et al., 2013). The main characteristics of seamless learning are shown as follows: (1) Encompassing formal and informal learning, (2) Encompassing personalized and social learning, (3) Across time, (4) Across locations, (5) Ubiquitous knowledge access, (6) Encompassing physical and digital worlds, (7) Combined use of multiple device types, (8) Seamless switching between multiple learning tasks, (9) Knowledge synthesis, (10) Encompassing multiple pedagogical or learning activity models (Wong & Looi, 2011).

One of its most important issues is how to bridge in-class and out-of-class language learning because this is inevitable in designing both in-school and out-of-school activities to link what students have learned in school with their daily life experiences and vice versa, particularly, to link what they have learned in their daily lives to their experiences in class. In the field of seamless language learning, majority of researchers reported that seamless learning is expected to enhance the learning effect and motivation. For example, Wong et al. (2014) reported a seamless learning system called MyCLOUD (My Chinese UbiquitOUs learning Days), which allows students to learn the Chinese language in both in-school and out-of-school learning spaces. Uosaki et al. (2010) reported a seamless learning system called SMALL System (Seamless Mobile-Assisted Language Learning support system) to support Japanese students who aimed to learn the English language in a formal and informal setting. Most of these studies focused on realizing a seamless learning environment at school or university. Once realized, the students’ learning logs have been accumulated into their server.
Therefore, we contend that learning efficacy can be enhanced by utilizing their accumulated learning logs. Presently, this aspect has attracted little attention. The research issues of learning analytics based on seamless learning environments are as follows:

- How can we utilize the learning logs accumulated in seamless language learning system?
- How can analytics bridge the gap between formal and informal learning?
- How can analytics enhance and support the students' learning experiences?

To address these issues, this study proposed a seamless visualization and analysis system called VASCORLL 2.0 (Visualization and Analysis System for Connecting Relationships of Learning Logs). VASCORLL 2.0 analyzes and visualizes learning logs accumulated in seamless language learning system. The system supports eBook-based learning and real-life learning by integrating the ubiquitous learning system called SCROLL with e-book system. To evaluate (1) whether VASCORLL 2.0 is effective in bridging the words learned in eBook to those learned in real-life, and (2) which social network centrality is the most effective to enhance the learning opportunities for seamless language learning, an experiment was conducted.

Literature review

Authentic learning and seamless language learning

Many empirical research have found that classroom-only learning is not conducive to enhance the learners’ communicative skills, such as listening and speaking, and sustaining their learning motivation. To provide effective language learning, it is necessary to consider in-class learning and out-of-class learning or authentic learning. Effective language learning is characterized by the active and constructive production of thoughtful linguistic artifacts in authentic learning settings (Ellis, 2000).

Authentic learning is either experimental learning or real-life learning. To bridge in-class learning and real-life language learning, most Seamless Language Learning (SLL) studies have been conducted from learning technologists’ perspective, which typically prioritizes the utilization of mobile affordances. In addition, these studies have been designed to introduce seamless learning environments into universities or schools (Wei, 2012; Ogata et al., 2015; Chai et al., 2016). However, they did not consider the learning analytics perspective for supporting seamless learning (Hwang et. al., 2017). In summary, SLL needs to consider seamless learning supports from learning analytics perspectives because the collected learning logs aren’t utilized to support learning. We believe that utilizing the collected learning logs lead to enhancing the quality of learning.

To support SLL from learning analytics perspectives, this study proposes seamless learning analytics to bridge eBook-based learning and real-life language learning by analyzing and visualizing the learning logs collected in both formal and informal settings. In the section entitled “SCROLL,” this paper describes the seamless language learning for supporting real-life language learning.

E-book learning system and e-book-based learning analytics

E-books represent a combinational figure in printed books and interactive computing technologies. Japanese government has announced their future policy to introduce “instead of” as a policy that introduces e-books in all K12 schools by 2020 (MEXT). The e-book policies of many countries only focus on introducing the technology of e-books into K12 schools (Fang, Liu & Hung, 2011; Shin, 2012). Some of the benefits of e-books have been identified, such as easily portable, easy search, highlighting, copy and paste, and quick updates.

Many previous studies of e-book have addressed above several benefits and enhancing opportunities for learners to interact with the learning contents. For example, Hwang et al. (2017) proposed an interactive e-book-based flipped learning approach to facilitate and bridge in-class learning and out-of-class learning. The contribution of their study is to bridge what learners have learned at home to the in-class activities by using their e-book system. They suggested that it is necessary to consider learning log analysis to better assist students and teachers, as their future works. However, little attention has been paid to visualizing and analyzing the e-book logs. Therefore, it is necessary to explore various analytics in this aspect.

This paper calls the visualizing and analyzing of e-book activity logs as “E-book-based Learning Analytics (ELA).” Some researchers at Kyushu University in Japan reported several analytics using a document viewer system called BookLooper (Ogata et al., 2015; Yin et al., 2014; Kiyota et al., 2016; Mouri et al., 2016a; Mouri et
The objectives of their studies are as follows: (1) improvement of learning materials, (2) analyzing learning patterns, (3) detecting the students’ comprehensive level, (4) predicting final grades, and (5) recommending e-books in accordance with personalization.

With consideration of the discussions of the ELA, we developed an e-book viewer system. In the section titled “E-book system for supporting language learning”, this paper describes the e-book system for supporting seamless language learning.

Previous work

E-book system for supporting seamless language learning

Figure 1 shows the interface of e-book library. The teachers or instructors create e-book contents using PowerPoint and Keynote prior to the class and use them in their courses. The uploaded e-book contents are converted to EPUB format and it is supported to access the contents by using smartphones and PCs. Figure 1 (Left) shows each directory in their course. For example, when the learner clicks the directory of the course named “Onomatopoeia class,” the learner can see the uploaded e-book contents in the directory. Figure 1 (Right) shows the files in the directories.

Figure 2 shows the e-book viewer interface and slide descriptions uploaded by the teachers. The learners can read the digital textbooks on their web browser anytime and anywhere. For example, when a learner clicks the memo button on the e-book viewer system, the learner can write a description concerning the target words as shown in Figure 2 (Right-top). In addition, the learner can search the page number corresponding to the target word in the e-book by clicking the search button as shown in Figure 2 (Right-bottom). Their operation logs, such as opening a book, zooming, bookmark, memo, searching words, highlighting important words, and page turning, are collected into the database server.
SCROLL

SCROLL project has started to support real-life language learning for international students since 2011. SCROLL aims to aid users to simply capture, review and reflect their learning logs, reuse and share the knowledge. To simplify the process of capturing the learning experience in their daily life, SCROLL provides a well-defined form to illustrate a learning log. It adopts an approach to share contents with other users based on a LORE (Log-Organize-Recall-Evaluate) model proposed by (Ogata et al., 2011). How the model supports each learning process is described below.

- Log: International students are likely to face some problems such as how to read, write and pronounce words in their daily life. They can save what they have learned with photo, such as location (latitude and longitude), learning place, and date and time of creation as a ULL (Ubiquitous Learning Log) as shown in Figure 3.

- Organize: When an international student adds a new ULL, SCROLL compares it with his past ULL and those of other users, categorize it and shows him related ULLs. By sharing ULLs as shown in Figure 4, past learning and current learning can be linked and their knowledge will be reorganized and reinforced.

- Recall: Learners are likely to forget what they have learned previously. It is necessary to support re-calling their past ULLs. During this learning process, the system support learners to recall what they have learned by using a quiz as shown in Figure 5 (Li et al., 2013; Ogata et al., 2014). The quiz are created automatically from uploaded ULLs. By answering the quiz, the learner’s knowledge will be enhanced.

- Evaluate: It is important to recognize what and how the learner has learned by analyzing the past ULLs, so that he or she can improve what and how to learn in the future. Mouri et al. (2014; 2015a; 2015b) developed an innovative visualization system that implemented Time-Map with network based graph theory to support this learning process. For example, when learners use the visualization system, they can reflect on what and how they have learned based on their past ULLs. It is expected that enhancing learning activities to share and reflect ULLs.

Figure 3. Add a ULL

Figure 4. A ULL
VASCROLL

Our previous VASCORLL could visualize and analyze learning logs accumulated in SCROLL to support real-life learning (Mouri et al., 2016c). For example, there is a ULL where an international student learned “fan” at the university in the past. It means “扇風機 (mechanical fan)” in Japanese. There is another ULL where another international student learned the same word, “fan” in a different context in the past. In this case it means “うちわ (Uchiwa is a round, flat paper fan with a wooden or plastic handle)” in Japanese. Even if the English word is the same, the meaning might be different if the context is different. By using VASCORLL, they can learn such relationships.

The results of an evaluation experiment indicated that VASCORLL was useful tool in detecting the correlations among learners, words and location in a ubiquitous learning environment. Furthermore, VASCORLL could increase learners’ learning opportunities and learners can apply their own experiences to different learning places.

However, the system did not consider learning analytics in the seamless learning in order to find central words bridging e-Book learning over real-life learning. Therefore, this study developed VASCORLL 2.0 based on the previous work.

Method

VASCORLL 2.0 framework

The aim of VASCORLL 2.0 is to support learners to apply what they have learned in class to their daily life experiences and vice versa, particularly, what they have learned in their daily life to e-Book learning. To bridge both e-Book learning and real-life learning, this study designed innovative visualization structures as shown in Figure 6: E-Book Learning Structure (ELS) and Real-life Learning Structure (RLS). ELS consists of three layers, which are called “eBook learner,” “Words learned through eBooks,” and “learning materials.”

- Words learned through eBook: The intermediate layer shows words that they have learned using e-book viewer interface.
- Learning materials: The Lowest layer shows learning materials uploaded by teachers or instructors.

To visualize the relationships among eBook learner, words learned through eBooks and learning materials, this paper visualizes the relationships based on network with directed graph. How our visualization method does connects relationships of each node? For example, if a learner learns and saves a newly learned word using e-book viewer interface, our visualization method will connect the learner’ node in the upper layer in the ELS to the word’ node in the intermediate layer in ELS. Moreover, the word’ node will connect to the learning material nodes in the lowest layer in ELS. By visualizing these links, teachers and students can grasp which e-book learning contents have the target word.
RLS includes three layers, which are called “Real-life learners,” “Words learned in a real-life,” and “Locations.”

- **Real-life learners:** The upper layer shows learners studying in an informal setting such as museums, restaurants and city halls using SCROLL.
- **Words learned in a real-life:** The intermediate layer shows words that they have learned in a real-life setting using SCROLL.
- **Locations:** The lowest layer shows contextual data such as location and place where they have learned in a real-life setting using SCROLL.

![Visualization structures](image)

**Figure 6.** Visualization structures in the seamless learning environments: E-Book Learning Structure (ELS) and Real-life Learning Structure (RLS)

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Formula (graph G:=(V,E) with V vertices and E edges)</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Degree</strong></td>
<td>$C_i^D = \frac{k_i}{N-1}$</td>
<td>Degree centrality is defined as the number of links incident upon a node. That is, it is the sum of each row in the adjacency matrix representing the network. N is the number of node and $k_i$ is the degree of the node i.</td>
</tr>
<tr>
<td><strong>Closeness</strong></td>
<td>$C_i^C = (L_i) = \frac{N-1}{\sum_{j \in \mathcal{E}, j \neq i} d_{ij}}$</td>
<td>Closeness centrality is that the distance of a node to all others in the network. $d_{ij}$ is the shortest path length between i and j, and $L_i$ is the average distance from I to all the other nodes.</td>
</tr>
<tr>
<td><strong>Betweenness</strong></td>
<td>$C_i^B = \frac{1}{(N-1)(N-2)} \sum_{j \in \mathcal{E}, j \neq i} \sum_{k \neq i, k \neq j} \frac{n_{jk}(i)}{n_{jk}}$</td>
<td>Betweenness centrality is that the number of shortest paths between any two nodes that pass via a given node. $n_{jk}$ is the number of the shortest path between j and k, and $n_{jk}(i)$ is the number of the shortest path between j and k that contains node i.</td>
</tr>
</tbody>
</table>

Based on visualization structures in the seamless learning environments described previously, this study presents the centralities using social network analysis as shown in Table 1. Degree, closeness and betweenness centralities are the fundamental measurement concepts for the social network analysis (Latora & Marchiori, 2004; Freeman, 1979). Particularly, we hypothesize that the betweenness centrality could bridge the gap between e-Book learning and real-life learning. For example, if an international student learns word “natto” on e-book content in a formal setting, it might be applied to various learning places such as supermarkets, shopping malls, and restaurants in an informal setting.
However, it is difficult for him/her to know whether it can be learned in other learning environments or where it can be learned. In addition, it is difficult for learners to find which words were most frequently learned by learners in a variety of learning environments. These words could play the most important role to bridge over eBook learning and real-life learning to realize the seamless learning environments.

**Implementation**

**VASCORLL 2.0 interface**

Majority of network graph studies have focused on advantages such as good-quality results, flexibility, simplicity, and interactivity. For example, a network layout called “force-directed” uses the force vector algorithm proposed in the Gephi software, which is appreciated for its simplicity and for the readability of the network that helps in the visualization (Mathieu et al., 2014; Noack, 2009; Fruchterman & Reingold, 1991). A network layout called “Yifan Hu multilevel” uses a very fast algorithm to reduce complexity (Hu & Scolt, 2001). The repulsive forces on one node from a cluster of distant nodes are approximated by a Barners-Hut calculation, which treats them as one super-node (Barners & Hut, 1986).

![Figure 7: VASCORLL 2.0 interface and learning scenarios](image)
Moreover, Mouri et al. (2015a) proposed Ubiquitous Learning Graph (ULG), which is divided into four areas: top-left, top-right, bottom-left, and bottom-right. In their evaluation experiment, they reported that it is important to establish their nodes’ position on the network graph for readability and ease-of-use when visualizing the relationships in the real-world language learning. Considering these points, we developed a network layout called “Seamless Learning Graph (SLG),” which is divided into six areas as shown in Figure 7 (Top): upper-left (eBook learners), center-left (Words learned through eBook), bottom-left (Learning materials), upper-right (Real-life learners), center-right (Words learned in a real-life), and bottom-right (Locations). These areas represent the layers as shown in Figure 6.

The interface implements three centralities based on social network analysis. Buttons (1) to (3) in Figure 7 show each centrality. By clicking them, VASCORLL 2.0 will automatically visualize and analyze all learning logs accumulated in the e-book system with SCROLL. The node size is based on the numerical value of each centrality. Figure 7 (Middle) shows the enlarged graph in both words learned through eBook and in a real-life areas. There are two learning scenarios by utilizing the results of visualization, which are called “Learning via Words learned through eBook” and “Learning via words learned in a real-life” as shown in Figure 7 (Bottom).

- Learning via words learned through eBook: As shown in Figure 7 (Middle), the word “natto” is the biggest size in the words learned through eBook area. By clicking it, the system moves to the page where the word “natto” appears. In this manner, learners can grasp which eBook and which page includes it.
- Learning via words learned in a real-life: After learning “natto” in the eBook contents, learners can find “natto” in the words learned in the real-life areas. By clicking it, the system moves to the ULLs (“natto” pages of SCROLL) learned in the real-life setting. Unlike the above learning method (1), by utilizing the ULL, they can share and learn other learners’ learning experiences (not only words but also time, location and place information) that cannot be learned in the formal setting.

Evaluation

Participants

Twenty international students who are studying at the University of Tokushima and Osaka in Japan participated in the evaluation experiments. The students were from China, Malaysia, Thailand, and Mongolia and aged from 21 to 36 years old. Their length of stay in Japan ranged from 1 month to 5 years. The evaluation experiment was designed to evaluate the following three points:

- Whether VASCORLL 2.0 can increase the participants’ learning opportunities (“Learning opportunities” denote that the number of ULLs that the learner uploaded to the system during the evaluation period”).
- Whether VASCORLL 2.0 would be benefit in terms of usability in finding important words in the seamless learning environment.
- Which centrality is effective in supporting learning in the seamless learning environment?

Method

A Japanese instructor uploaded e-book contents to the server prior to his/her class. The uploaded e-book contents were categorized according to JLPT (Japanese Language Proficiency Test). At the beginning of the first week, the instructor had a briefing session on how to use e-Book system with SCROLL since it was their first time to use it. They practiced using the e-book system with SCROLL for one week before the information of VASCORLL 2.0. Based on the uploaded ULLs during the first week, the participants were divided into two groups as even as possible in terms of the keenness of language learning: Group A (Experimental group) and Group B (Control group). Figure 8 shows the experimental procedure.

Table 2 shows the number of ULLs that the participants uploaded during the first week before using VASCORLL 2.0. Group A participants uploaded 143 ULLs and Group B participants uploaded 149 ULLs to the system. The means and standard deviations were 14.3 and 6.78 for Group A, and 14.79 and 6.51 for Group B. The t-test shows that there was no significant different between the two groups ($t = 0.201, p > 0.05$). This result indicates that the participants of the two groups have learning opportunities before using VASCORLL 2.0. Then, the instructors introduced how to use VASCORLL 2.0.
Figure 8. Experimental procedure

Table 2. Number of uploaded ULLs in the first week (practice period)

<table>
<thead>
<tr>
<th>Group</th>
<th>Number of participants</th>
<th>Number of ULLs</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>10</td>
<td>143</td>
<td>14.3</td>
<td>6.78</td>
<td>0.201</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Group B</td>
<td>10</td>
<td>149</td>
<td>14.9</td>
<td>6.51</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Group A consisted of 5 Chinese, 4 Mongolians and 1 Malaysian. Group B consisted of 3 Chinese, 5 Mongolians and 2 Thais. Group A learned words in their daily lives and words in the e-book contents using the e-book system with SCROLL and VASCORLL 2.0. Group B learned words in their daily lives and words in the e-book contents without VASCORLL 2.0. The participants used their own smart-phones (iPhone 4s or Android device) to upload their ULLs in formal settings and informal settings anytime and anywhere. The mobile devices used in the evaluation experiment were three iPhone 4s, fourteen iPhone 5s, and three Samsung Galaxy Note 3s. In the second week, they were required to conduct real-life learning activities described previously.

During the third week, both Groups A and B participants evaluated each centrality based on social network analysis during real-life learning activities. The reason for integrating both Group A and B is to evaluate three centralities with enough number of participants in terms of easiness to find words bridging eBook and real-life learning. They learned words using three algorithms: degree, closeness and betweenness. They were asked to use the prearranged one centrality (e.g., participants first had to use degree centrality for two days). After the evaluation experiment, the participants were asked to complete five-point-scale questionnaires to evaluate the system performance and usability, as well as the user-friendliness of understanding the contents and finding ULLs using each centrality in VASCORLL 2.0.

Result and discussion

To examine the increase of participants’ learning opportunities by our proposed VASCORLL 2.0, we compared the number of the uploaded ULLs of Group A with that of Group B. Two-way ANOVA was applied to understand the effect on the learning opportunities of the different groups (with and without VASCORLL) and the time of measurements (1st week and 2nd week), (Liu et al., 2014). Table 3 shows the means and standard deviations for experimental and control groups in the 1st and 2nd week.

The result of repeated measures analysis showed that interaction effect between group and time of measurements was significant, (F = 4.11, p ≤ .1). In other words, VASCORLL 2.0 was able to increase their learning opportunities. Figure 9 shows the number of their uploaded ULLs from the first to the second week. The
horizontal axis represents the first and second week, and the vertical axis represents the number of their uploaded ULLs. The number of uploaded ULLs of Group A increased better than that of Group B. These results imply that VASCORLL 2.0 plays an important role to increase the number of their uploaded ULLs.

Table 3. Number of uploaded ULLs in the second week

<table>
<thead>
<tr>
<th></th>
<th>1st week</th>
<th>2nd week</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>14.3 (6.78)</td>
<td>18.9 (6.41)</td>
<td>4.11</td>
<td>.08</td>
</tr>
<tr>
<td>Group B</td>
<td>14.9 (6.51)</td>
<td>127 (6.75)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 9. Comparison between the number of the uploaded logs during the 1st and 2nd week

Further, a one-way ANOVA was adopted to compare the high-motivated group and low-motivated group. The participants in the top 50% in terms of the number of the uploaded ULLs were categorized as high-motivated, while the others were as low-motivated. Table 4 shows the statistical results. A significance was found in the low-motivated group, implying the low-motivated participants were able to increase the number of their uploaded ULLs using VASCORLL 2.0. Therefore, it was detected that VASCORLL 2.0 was a useful tool to increase their learning opportunities.

Table 4. Summary of analysis of the number of uploaded ULLs in different learning motivation

<table>
<thead>
<tr>
<th>Learning motivation</th>
<th>Group</th>
<th>Number of participants</th>
<th>Number of ULLs</th>
<th>Mean</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Group A</td>
<td>5</td>
<td>119</td>
<td>23.8</td>
<td>3.88</td>
<td>&gt; .05</td>
</tr>
<tr>
<td></td>
<td>Group B</td>
<td>5</td>
<td>80</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>Group A</td>
<td>5</td>
<td>70</td>
<td>14</td>
<td>5.31</td>
<td>&lt; .05</td>
</tr>
<tr>
<td></td>
<td>Group B</td>
<td>5</td>
<td>47</td>
<td>9.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The results of the five-point-scale questionnaires are presented in Table 5 (Best: 5, Worst: 1).

Table 5. Result of the five-point-scale questionnaire for evaluating the usability of the system (Group A)

<table>
<thead>
<tr>
<th>Question</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1. Were you able to connect words learned through eBook to those learned in a real-life by using VASCORLL 2.0?</td>
<td>3.7</td>
<td>0.82</td>
</tr>
<tr>
<td>Q2. Were you able to connect words learned in a real-life to the words learned through eBook by using VASCORLL 2.0?</td>
<td>3.9</td>
<td>0.99</td>
</tr>
<tr>
<td>Q3. Were you able to learn and find the relationship between words and places by using VASCORLL 2.0?</td>
<td>3.4</td>
<td>0.96</td>
</tr>
<tr>
<td>Q4. Were you able to find your newly learned words used in other eBook contents by using VASCORLL 2.0?</td>
<td>3.5</td>
<td>0.87</td>
</tr>
<tr>
<td>Q5. Was VASCORLL 2.0 easy to use?</td>
<td>2.6</td>
<td>1.23</td>
</tr>
</tbody>
</table>

Q1 asks whether the participants were able to find that the words they learned through eBooks were connected to the words that other learners learned in a real-life learning setting. Similarly, Q2 asks whether they were able to find that the words they have learned in a real-life learning setting were connected to the words that others have learned through eBooks. The results of Q1 and Q2 revealed that the participants found the words they learned
through eBooks were connected to the words learned in a real-life, vice versa. For example, some participants learned the Japanese word “納豆 (natto)” through e-book contents. By uploading “natto” to the system, the system showed them that other participants have learned it at the shopping mall and supermarkets. In this manner, VASCORLL 2.0 was able to connect the words between eBook learning and real-life learning.

Q3 asks whether the participants were able to find that their newly learned words were connected to the places where other learners learned the identical word by using VASCORLL 2.0. For example, when a participant learned the Japanese word, “料理 (Cuisine)” in eBook, he/she could find that it was connected to the experiences of others at places such as schools and restaurants. By sharing the authentic experiences, VASCORLL 2.0 enabled them to experience indirectly what other people experienced, which connected their eBook learning to the real-life learning.

Q4 asks whether the participants were able to find their newly learned words in other e-book contents. For example, when a participant learned “使用 (Use)” in an e-book material titled “Japanese Learning Beginner Vol.1”, the system connected it to other e-book materials such as “Japanese Learning Beginner Vol. 2” and “Onomatopoeia Japanese Learning Vol. 1.” In this manner, they could learn that it was a frequently used word in the Japanese language.

Q5 asks whether VASCORLL 2.0 was easy to use. They were asked to evaluate the usability in terms of operability and readability of the visualized graph. The response shows that many participants felt that VASCORLL 2.0 was not easy to use. They were asked to give comments regarding this problem, and the negative comments are as follows:

- The speed of visualizing and analyzing logs in the system is too slow (about 20~30 sec.).
- If visualizing logs using a mobile device, it is hard to read the nodes because of very small screen size. However, if logs are visualized on a personal computer, they become very easy to read.
- It was a little bit difficult to understand how to use the system.

From comments (1) and (2), the participants would suggest that the system developers need to improve the functionality in accordance with their mobile device and system’ speed in visualizing a large number of logs. Comment (3) shows that even though we explained the usage of VASCORLL 2.0 prior to the evaluation experiment, some participants did not understand fully how to use it. Thus, our next evaluation should be more carefully planned.

Table 6 shows the result of the five-point-scale questionnaire for evaluating each centrality (degree, closeness, betweenness). Moreover, the participants were asked questions such as “Which centrality is the easiest in understanding or finding central words?” and “Which centrality is the most effective for learning” in order to evaluate each centrality in the seamless learning environment.

Q1-Q3 asked whether the participants were able to understand and find central words using fundamental the social network analysis: degree centrality, closeness centrality, and betweenness centrality. From the results of the questionnaire, most participants preferred to learn and find central words using betweenness centrality because the mean score of the Q3 is the highest.

<table>
<thead>
<tr>
<th>Question</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1. Was the degree centrality easy to understand in finding central words in the network graph?</td>
<td>3</td>
<td>0.91</td>
</tr>
<tr>
<td>Q2. Was the closeness centrality easy to understand in finding central words in the network graph?</td>
<td>2.7</td>
<td>1.08</td>
</tr>
<tr>
<td>Q3. Was the betweenness centrality easy to understand in finding central words in the network graph?</td>
<td>3.7</td>
<td>0.86</td>
</tr>
</tbody>
</table>

To find the most effective centrality to learn central words, we interviewed the participants to compare betweenness centrality with other centrality.

**Degree centrality versus betweenness centrality**

Degree centrality enabled the participants to find merely nodes that have many links. Two participants selected the centrality in terms of usability and effectiveness for learning because it is simple and easy to understand the...
characteristics as shown in Figure 10. However, most participants commented that it was difficult to find words bridging eBook learning over real-life learning because the size of eBook or real-life learner nodes became larger than those of nodes.

Closeness centrality versus Betweenness centrality

When comparing closeness centrality with the betweenness centrality, the closeness centrality was not useful to find central words in the seamless learning environment. There was no numerical value difference between words learned through eBook and words learned in a real-life, so that the participants could not find central words. Therefore, this paper concluded that the closeness centrality was not a useful centrality in finding central words in the seamless learning environment using our visualization and analysis method.

As shown in Figure 10, majority of the participants preferred to use betweenness centrality than other centrality. We asked them “why you preferred to use the betweenness centrality than the other centrality.” Their feedbacks are as follows:

- “The betweenness centrality was very good because it was easy to find words in my e-book contents linking to words in a real life”
- “It was easy to understand. And I learned some words. Then, it recommended some useful words to me (e.g., The different sizes of the nodes (large and small) and color coding such as green or yellow noes were easy to recognize).”

The betweenness centrality was the best centrality of all in terms of easiness to find words bridging eBook and real-life learning. In addition, we compared the betweenness centrality with other centrality. Fifteen participants answered that the betweenness centrality is helpful in finding central words in the seamless learning environment. Seventeen participants answered that it worked effectively in language learning.

Conclusion

This paper described a system called VASCORLL 2.0 for visualizing and analyzing learning logs collected in the seamless language learning environment in order to bridge the gap between eBook learning and real-life learning. VASCORLL 2.0 works on cyber-physical setting to link learners in the real world and learning logs that are accumulated in the cyber spaces using e-book system with the ubiquitous learning system called SCROLL. The e-book system with SCROLL enabled international students to learn through two learning activities: e-book-based learning activity and authentic learning activity.

Using these learning activities, we proposed the visualization and analysis methods based on graph theory, social network analysis and graph drawing algorithms in order to find pivotal words in the seamless learning environment. Two types of three-layer structures called ELS and RLS were adapted as the visualization methods. In this manner, teachers and students could easily grasp words bridging between words in ELS and RLS.
addition, this paper evaluated whether they were able to find the most pivotal words on the network graph using each centrality based on social network analysis.

The evaluation was conducted after the implementation of VASCORLL 2.0. A questionnaire with a five-point-scale conducted after the evaluation showed that VASCORLL 2.0 was a useful tool in finding central words bridging eBook learning and real-life learning. The result of questionnaires for evaluating each centrality showed that the most effective centrality for learning was betweenness centrality. Therefore, we concluded that the betweenness centrality is the most important centrality in the seamless learning environment.

VASCORLL 2.0 will be evaluated repeatedly, with the improved processing speed of visualizing and analyzing learning logs improved. In addition, our future works include applying VASCORLL 2.0 to other application domains such as math, physics, and science education, and long-term evaluations with a sufficient number of participants.

Acknowledgements

The part of this research work was supported by the Grant in-Aid for Scientific Research No. 16H06304 and No. 17K12947 from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) in Japan.

References


How Competition in a Game-based Science Learning Environment Influences Students’ Learning Achievement, Flow Experience, and Learning Behavioral Patterns

Ching-Huei Chen*, Jun-Han Liu and Wen-Chuan Shou
Department of Industrial Education and Technology, National Changhua University of Education, Taiwan, R.O.C. // chhchen@cc.ncue.edu.tw // junhan657@gmail.com // wen01306132@gmail.com

*Corresponding author

ABSTRACT
Although educational games have become prevalent in recent research, only a limited number of studies have considered learners’ learning behaviors while playing a science problem-solving game. Introducing a competitive element to game-based learning is promising; however, research has produced ambiguous results, indicating that more studies should investigate its pros and cons of competition. A total of 57 seventh-grade students participated in the study and were assigned to two conditions: competition or non-competition. Results revealed that students in the non-competition condition performed significantly better on the learning achievement test than those in the competition condition. With regard to the flow experience, no significant differences were found between the two conditions. The results of learning behavioral analyses revealed that, while both conditions resulted in students acquiring through means-ends strategies, students in the non-competition condition tended to read the instructions carefully and repeatedly sought additional supports to help themselves advance their conceptual understanding. These findings, when examined in light of previous research, call into question other types of competition in promoting engagement and supporting learning.

Keywords
Game-based science learning, Competition, Learning performance, Flow experience, Learning behavioral patterns

Introduction
In general, a game is organized play with a set of rules and obstacle apparatuses (e.g., Ke, 2016; Klopfer et al., 2009). Characteristics of a game often include interactivity, rules and constraints, goal(s), competition, and feedback (Wouters & van Oostendorp, 2013). In the last decades, game-based learning (GBL) has become a prevalent instructional approach across various disciplines. Moreover, GBL is situated in an interactive environment based on a set of agreed-upon directions and constraints (Garris, Ahlers, & Driskell, 2002) and is aimed at solving a clear goal that is often set by a challenge (Malone & Lepper, 1987). GBL provides constant feedback, either as a score or as changes in the game world, to enable players to monitor their progress toward the goal (Prensky, 2001). While many efforts have been pouring into the study of GBL, evidence of its effectiveness on learning has yet to reach consensus across the different disciplines. Mayer (2011) suggested that game research should take further approaches in value-added and cognitive consequences to understand how specific game feature fosters learning and motivation as well as to determine what students learn from games.

Therefore, in this study, we focused on the distinct features that constitute an educational game in order to define its effectiveness (Aldrich, 2005; Vandercruysse et al., 2012). Competition, one of the dispensable game features, has attracted the attention of many researchers because it is so multifaceted in nature and diverse in form. Some researchers have considered competition a motivational trigger in a game because it stimulates student engagement and persistence in the learning activity (Kollöffel & De Jong, 2016; Malone & Lepper, 1987). Yet others warn that overemphasizing the competition aspect may induce negative emotional effects and ultimately distract students from learning content and seeking support (Cheng, Wu, Liao, & Chan, 2009; Van Eck & Dempsey, 2002). In light of this, the competition element in GBL has inconsistent findings, and contextual considerations require further investigation. In this study, we set out to examine the effect of competition in a game-based science learning (GBSL) environment and to determine the differences in student learning behavioral patterns when competition as an element. This study specifically attempted to answer the following research questions:

- Does the competition mode in GBSL affect students’ learning achievement when compared to the non-competition mode?
- Does the competition mode in GBSL affect students’ flow experience when compared to the non-competition mode?
- Does engaging in the competition mode in GBSL affect students’ learning behavioral patterns differently than those who engage in the non-competition mode?
Review of relevant literature

Game-based science learning (GBL)

Consensus is growing among science educators and practitioners regarding the importance of designing personally relevant experiences to foster lifelong learning in science (Eylon, 2000). Studies have shown that GBL provides a learner-centered context that enhances learning (Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012). Particularly, indulgence in GBL presents a unique opportunity for science learning since the subject of science is universally acknowledged to be complicated and challenging. Cheng and her colleagues (2016) found that GBL (learning by gaming) facilitated student science learning achievement. Moreover, GBL is often believed to have the potential to facilitate both the cognitive and affective/motivational processes of learning (Ke, 2016; Wouters, van Nimwegen, van Oostendorp, & van der Spek, 2013). A meta-analytic examination on the instructional effectiveness of GBL also found a significant improvement in learning achievement (Sitzmann, 2011). While GBL offers many educational advantages that can address learning needs, meta-analyses of GBL have not reached consistent findings about the impact of GBL in the context of science education. Some studies have found that games have a significant impact on science learning, while others identified insignificant relationships between the two (Vogel et al., 2006; Wouters & van Oostendorp, 2013). Additionally, several researchers have suggested that different elements of the GBL environment should be separately examined before a conclusion on the effectiveness of an educational game is reached (e.g., Young et al., 2012). Competition, amongst various elements that could have affected learning behaviors and flow experience, is multifaceted in nature and diverse in form. Next, we summarize the literature on the competition element in GBL.

Competition in GBL

Competition is an important ingredient of GBL. Game developers have always maintained that competition allows for GBL activities that are intrinsically game-like (e.g., Prensky, 2001). Of course, there have been fluctuations in emphasis, and much has changed through the years. Ideally, when competing with others, learners must work harder; as a result, all students improve their knowledge, allowing the group to progress. Without competition, only the best of the class would improve his or her knowledge. In GBL, Cheng and his colleagues (2009) considered competition a well-structured learning activity with the potential to increase learners’ attention and generate excitement. As a result, researchers have regarded competition as a useful technique to motivate individuals to learn (Burguillo, 2010; Julian & Perry, 1967; Malone & Lepper, 1987; Nemerow, 1996). For instance, van Eck and Dempsey (2002) reported that competition may motivate students extrinsically, resulting in students putting forth more effort on current tasks. Many studies have found positive effects of competition in GBL such as enhanced learning and motivation, a higher tendency to make risky plays, and decreased cognitive load (Admiraal, Huizenga, Akkerman, & Dam, 2011; Cagiltay, Ozcelik, & Ozcelik, 2015; Foo et al., 2017; Hwang & Chang, 2015).

However, competition may not always be prominent in GBL. Vander cruysse, Vandewaetere, Cornillie, and Clarebout (2013) found that competition had no effect on learning outcomes and/or motivation. Furthermore, competition may weaken a student’s intrinsic motivation to learn the educational content because of the focus on winning (Van Eck & Dempsey, 2002). In line with Van Eck and Dempsey’s (2002) argument, competition may increase hostility between students, and such social comparison can negatively affect a student’s self-efficacy beliefs, motivation, and performance (Bandura & Locke, 2003). The abovementioned studies have generally shown that competition may inhibit metacognitive skills, attention, and elaboration and may also create an affective state of anxiety that may be detrimental to learning.

After reviewing these studies, we concluded that the inconsistent findings regarding competition in GBL may be due to differences in the forms of competition and the learning content itself. Competition in GBL takes many forms; for example, players may compete with themselves, with the game system, with other individual players, with other teams, or a combination of these to achieve game objectives (Alessi & Trollip, 2001; Fisher, 1976; Yu, 2003). In this study, we operationalized competition by allowing students to compare their own performance in GBL against that of their counterparts. Their learning content was science, which has a history of competition in traditional science fairs. Cheng and her colleagues (2009) pointed out that less-able students in science may be discouraged by the persistent winning of more-able students and thus, reduce their own effort. In other words, the same competition may result in different levels of engagement among different students. While the competition is assumed to be the key element in a game to foster motivation, there is limited research that addresses the empirical effectiveness of competition on learners’ affections. Considering the multifaceted nature
of competition, this study intended to investigate whether competition as an element in a GBL environment affects a student’s flow experience and triggers different learning behaviors when compared to a non-competitive environment.

**Flow experiences in GBL**

Over the past decades, researchers have been increasingly interested in understanding student motivation and finding ways to predict and improve it. According to self-determination theory, motivation can be affected by levels of control (Ryan et al., 2006), which implies that when learners make their own decisions in the learning process, they are more likely to be motivated. In the context of GBL, researchers have often considered motivation a key variable that can affect a learner’s motivational appeal (Garris et al., 2002; Malone, 1981). Ke (2009) claimed that, generally, instructional computer games seem to facilitate motivation across different learner groups and learning situations. Therefore, in this study, we examined one of the key characteristics of motivation, flow, which has been found to significantly predict perceived learning and enjoyment (Barzilai & Blau, 2014).

Flow is a state of mind in which a person is completely involved in an activity for its own sake. Many researchers have explored student flow experience in the context of GBL (e.g., Kiili, 2005; Sun, Kuo, Hou, & Lin, 2017). Admiraal, Huizenga, Akkerman, and Dam (2011) found that when experience flow in their game activities, that it has an effect on their game performance, but not on their learning outcome. Admiraal et al. (2011) found that distractive activities and being occupied with competition between teams affected the learning outcome of students. Their results revealed that the less students were distracted from the game and the more they were engaged in group competition, the more they learned about the medieval history of Amsterdam (Admiraal et al., 2011). Csikszentmihalyi (1991) stated that educators create flow by screening out distractions and making concentration possible. While game elements, such as challenge and complexity in GBL environment, were found to affect flow experiences of the students (Inal & Cagiltay, 2007), research addressing the effect of competition on the flow experience seems scarce. Presumably, while competition increases learners’ engagement (e.g., Malone & Lepper, 1987), there can also be cases where learners may be distracted or lack engagement when witnessing the persistent winning of other students (Cheng, Wu, Liao, & Chan, 2009). Therefore, in this study, flow is measured as an outcome variable that plays a motivational role in the learning process.

**Methodology**

**Design and participants**

A quasi-experimental study, supplemented by behavioral logs, was employed to investigate the three research questions. This study applied both quantitative and qualitative methods for the purpose of seeking triangulation of the results from different data sources and examining overlapping and different facets of a phenomenon. Participants in this study were 57 students (13-14 years old) who were recruited from two seventh-grade classes at a junior high school in Taiwan. One class, with 29 students (13 males and 16 females), was in the competition mode of GBL (“the competition condition”), while the other class, with 28 students (15 males and 13 females), was in the control group (“the non-competition condition”). A knowledge pre-test was administered and showed no differences among participants in terms of what they would be learning. Further, students reported equivalent experiences in learning through game playing.

**Game-based science learning environment (SumMagic)**

A central component of our study was a game called SumMagic, which was developed based on the theories of how we learn (e.g., Bransford, Brown, & Cocking, 2000). The topics for the learning subject were primarily in science, specifically, time, distance, position, and velocity. Learning objectives included: (a) to observe the timeline of an object’s movement, (b) to identify an object’s moving distance, (c) to define speed as a scalar quantity that involves a magnitude, and (d) to calculate an object’s speed from its moving distance and time data.

Students in SumMagic are required to learn independently using observable variables to analyze, operate, and finally rationalize the phenomenon and conclude the findings. The main functions covered in the game can be classified into four main components: (1) sources of game problem, (2) manipulative tools, (3) observable
interpretation, and (4) report results. Sources of game problem reflected the problem representation, in which students needed to utilize manipulative tools for clarifying the problem and the means to solve it. Based on prior problem-solving research (Bransford & Stein, 1984; Chi & Glaser, 1985), the representation of a problem consists of the solver’s interpretation of the problem, which, follows the connection of existing knowledge to form an integrated representation. During the learning process, students were required to interpret information about the problem by constructing the problem space and actively manipulating and testing before triggering particular solution processes. According to information-processing theories, if appropriate schema cannot be activated, the solver goes back to an earlier stage and redefines the problem or uses another method to solve the problem, so called means-ends analysis (Gick, 1986). Figure 1 provides the scenarios for SumMagic. This study evaluated competition as a specific game element. As shown in Figure 2, after the gameplay, students in the competition condition could see a ranking board that showed their own ranking as well as their counterpart’s ranking in score and time spent on completing the game. Conversely, students in the non-competition condition did not have access to the ranking board.

Figure 1. Game scenarios for SumMagic
Instruments

The data gathered for subsequent data analyses were learning achievement tests regarding time, distance, position, and velocity; surveys on flow experience; and learning behavioral logs. The learning achievement test was designed to assess the participants’ topic-specific knowledge. It consisted of 10 multiple-choice questions. The test was developed and validated by two experienced teachers who determined that the questions were well-represented and aligned with the content being taught. The reliability of the learning achievement test was .87.

The flow experience measurement was adopted from Killi (2006) and measured four key flow dimensions: concentration, loss of self-consciousness, transformation of time, and autotelic experience (Csikszentmihalyi, 1991). Previously, a Chinese version of Killi’s approach was used by Hou and Li (2014) and found to obtain reasonable reliability levels. Respondents completed a total of 12 questions using a 5-point Likert-type response format (5 = agree, 1 = disagree). Samples of questions were “My attention was focused entirely on playing the game” and “I really enjoyed the playing experience.” There were two additional open-ended questions that enabled students to describe whether they had experienced flow before and the factors that may interrupt their flow experiences. The reliability of flow experience measurement for this study was .92.

<table>
<thead>
<tr>
<th>Code</th>
<th>Learning behavioral dimension</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC</td>
<td>Speed change</td>
<td>Learners can change the speed level to complete the learning goals.</td>
</tr>
<tr>
<td>TC</td>
<td>Timer control</td>
<td>Learners can click on the timer to observe the changes in time while the objects move in distance and time.</td>
</tr>
<tr>
<td>RV</td>
<td>Ruler view</td>
<td>When scrolling over the moving object, students can use the ruler to observe its exact position.</td>
</tr>
<tr>
<td>BR</td>
<td>Board reading</td>
<td>Information about the goal, operation, and factual knowledge for completing the game.</td>
</tr>
<tr>
<td>RC</td>
<td>Role click</td>
<td>Learners can click on the non-player character (NPC) and answer the NPC’s questions.</td>
</tr>
<tr>
<td>NU</td>
<td>Note using</td>
<td>Learners can use the provided painter to take notes or use the calculator tool.</td>
</tr>
<tr>
<td>HU</td>
<td>Hint using</td>
<td>A list of prompts is provided for learners to guide their problem-solving processes.</td>
</tr>
<tr>
<td>GP</td>
<td>Game process</td>
<td>When learners complete the learning goals, the game will direct them to the next (more advanced) level of the game.</td>
</tr>
</tbody>
</table>

In addition to learning achievement tests and flow experience surveys, this study collected students’ gaming behavioral logs. The game automatically stored each student progressive game behaviors in the database. The content experts first reviewed the game and identified eight behavioral dimensions, which are defined in Table 1.
These behavioral dimensions provided a comprehensive view of how students performed and progressed within the game environment. These dimensions were also keys to solving the game problem and gaining important conceptual knowledge. For example, in order to understand the relationship between time, distance, and velocity, students needed to manipulate the time control (TC) and observe position change (RV) to experience the changes in velocity (SC). If the obtained velocity did not meet the requirements (RC), students then continually switched to other segments. Figures 3–7 exemplify how different dimensions were presented in the game.

![Figure 3. Screenshot of Board Reading (BR) interface](image)

![Figure 4. Screenshot of game interface (including SC, RV, and TC)](image)
Figure 5. Screenshot of game interface (including HU and RC)

Figure 6. Screenshot of game interface (including NU)
Procedure

The final analysis included only those students who were present for all phases of the study and for whom completed data were obtained. The study took place in the computer classroom with a teacher and two researchers present throughout. On the day of the study, the participants arriving in the computer classroom were seated in front of computers, introduced to the research team, informed of the general purpose of the study, and given a description of the procedures. After this orientation, students had approximately 20 minutes to individually complete an online pre-test on learning achievement. Upon completion of the pre-test, students were instructed to login to the game using a pre-assigned username and password. Students individually played the game at their own pace for approximately 40 minutes, and their gaming behaviors were logged and coded automatically through the computer system. The game ended once the students completed the challenges successfully. Upon the completion of the game play, students were instructed to complete a post-test and a flow experience survey.

Data analysis

To assess learning achievement, analysis of covariance (ANCOVA) was conducted, and a significance level of $p < 0.05$ was adopted. The assessment of flow experience was followed using $t$-test. All the analyses were done with the Statistical Package for the Social Sciences (SPSS 11.0 for Windows). The lag sequential analysis of student behavioral patterns was performed because it enables researchers to find significant behavioral patterns and determine whether the sequential relationships between each behavior pattern reached statistical significance (Bakeman & Gottman, 1997).

Results

Learning achievement

One of the objectives of this study was to examine whether the competition or the non-competition mode in GBSL influenced students’ learning achievement tests. ANCOVA was carried out to exclude the difference between the prior knowledge of the two groups by using the pre-test scores as the covariate and the post-test scores as dependent variables. The homogeneity test result showed that the post-test scores of the two groups were homogeneous ($F = 0.40, p = 8.42 > 0.05$), implying that ANCOVA could be applied. Table 2 summarizes the ANCOVA results in which the adjusted mean values of the post-test achievement scores were 8.57 for the competition group and 9.40 for the non-competition group; moreover, a significant difference was found between the two groups ($F = 7.40, p < 0.05, ES = .12$, observed power = .90), implying that there was a significant difference in the learning achievements of the competition and non-competition conditions.
Table 2. Descriptive statistics for the learning achievement post-test results

<table>
<thead>
<tr>
<th>Conditions</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Adjusted mean</th>
<th>Std. error</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Competition</td>
<td>29</td>
<td>8.66</td>
<td>1.82</td>
<td>8.57</td>
<td>.21</td>
<td>7.42*</td>
</tr>
<tr>
<td>Non-competition</td>
<td>28</td>
<td>9.32</td>
<td>0.77</td>
<td>9.40</td>
<td>.22</td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .05.

Flow experience

Table 3 shows the t-test result of the flow experience ratings of the two conditions. The means and standard deviations were 3.62 and 0.75 for the competition condition, and 3.48 and 0.79 for the non-competition condition. The t-test result showed no significant difference between the two groups (t = -.71, p > .05), indicating that the two groups of students reported equivalent flow experience after participating in the learning activity.

Table 3. t-test result of the flow experience for the two groups

<table>
<thead>
<tr>
<th>Conditions</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Competition</td>
<td>29</td>
<td>3.62</td>
<td>0.75</td>
<td>-.71</td>
</tr>
<tr>
<td>Non-competition</td>
<td>28</td>
<td>3.48</td>
<td>0.79</td>
<td></td>
</tr>
</tbody>
</table>

Learning behavioral patterns in different conditions

We collected a total of 13,921 behavioral codes (competition condition: 6,879, non-competition condition: 7,042). In terms of dimensional difference, Table 4 shows the frequency distribution of their learning behaviors across various dimensions.

Table 4. Summary of learner behaviors by frequency

<table>
<thead>
<tr>
<th>Conditions</th>
<th>BR</th>
<th>RC</th>
<th>GP</th>
<th>TC</th>
<th>RV</th>
<th>NU</th>
<th>SC</th>
<th>HU</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Competition</td>
<td>115</td>
<td>563</td>
<td>21</td>
<td>1549</td>
<td>2750</td>
<td>87</td>
<td>1654</td>
<td>140</td>
<td>6879</td>
</tr>
<tr>
<td></td>
<td>(2%)</td>
<td>(8%)</td>
<td>(1%)</td>
<td>(22%)</td>
<td>(39%)</td>
<td>(2%)</td>
<td>(24%)</td>
<td>(2%)</td>
<td></td>
</tr>
<tr>
<td>Non-competition</td>
<td>110</td>
<td>408</td>
<td>22</td>
<td>1441</td>
<td>2944</td>
<td>56</td>
<td>1664</td>
<td>397</td>
<td>7042</td>
</tr>
<tr>
<td></td>
<td>(2%)</td>
<td>(6%)</td>
<td>(1%)</td>
<td>(20%)</td>
<td>(41%)</td>
<td>(1%)</td>
<td>(23%)</td>
<td>(6%)</td>
<td></td>
</tr>
</tbody>
</table>

We first performed sequential analysis on the competition condition and adjusted the residual table on the dimensions and found that the significant sequences for p < .05 were RV → RV (z = 8.38), SC → SC (z = 4.262), SC → TC (z = 10.75), TC → SC (z = 6.188), RC → RC (z = 7.34), and RC → NU (z = 1.92). Figure 8 shows the learning behavioral patterns chart, and the thickness of the lines represents the closeness of the sequential relations. The competition condition had the following characteristics: (a) students preferred repeatedly using the ruler view (RV) to make observations when achieving their learning goals; (b) students tended to click on speed change (SC) many times, indicating they were likely to use means-ends and guessing strategies to answer their learning goals; (c) when the selected speed change (SC) was incorrect, students tended to work backward to click on the timer control (TC) to start over; (d) students preferred to interact or seek help from a role click (RC) to proceed; and (e) students liked to take or leave notes (NU) after seeking help from the NPC.

Figure 8. The dimensional behavioral patterns chart of the competition condition

We then adjusted the residual table based on the dimensions of the non-competition condition, we discovered that the significant sequences for p < .05 were SC → TC (z = 7.604), TC → SC (z = 13.453), RV → RV (z = 9.588), RV → RC (z = 2.157), RC → RV (z = 2.447), BR → RC (z = 3.422), RC → HU (z = 4.001), RC → GP (z = 3.8), and HU → HU (z = 9.563). The non-competition group’s learning behavioral patterns chart is shown in Figure 9. The learning behaviors of the non-competition condition exhibited the following characteristics: (a)
students tended to click on speed change (SC), indicating that students were likely to use means-ends and guessing strategies to answer their learning goals; (b) students preferred repeatedly using ruler view (RV) to make observations while also seeking help from the NPC (RC), which likely completed the learning goals and advanced them to the more difficult level of the game; (c) after interacting with the NPC, students also clicked on hint using (HU) repeatedly to seek additional support from the game itself; and (d) students preferred to engage in board reading (BR) before making observations.

Figure 9. The dimensional behavioral patterns chart of the non-competition condition

Discussion

The purpose of this study was to examine whether the presence of competition in GBSL made any differences on student learning achievement, flow experience, and learning behavioral patterns. The results confirmed the expectation that secondary students can benefit from a GBSL (e.g., Cheng et al., 2016). When comparing students’ performance tests, this study found the main effect of non-competition on GBL was in line with previous findings from studies on the use of competition in GBL (e.g., Vandercruysse et al., 2013). Specifically, this study compared the competition versus non-competition aspect of the game. The learning achievement tests showed that students in the non-competition condition performed significantly better than those in the competition condition. There might be several explanations for the results, including peer pressure and time constraints. The absence of peer pressure allowed students to continually work on the game challenges. From the field observations, we found that students in the non-competition condition spent most of the time exploring the problem space and solutions, and that they utilized the support tools provided in the game environment.

Regarding the flow experience as the result of gameplay, no significant difference was found between the competition and non-competition conditions. The insignificance may be due to several reasons. First, to experience flow, students should be immersed in the task for a longer or extended period of time. Second, to experience flow, extraneous distractions should be diminished so that concentration can be facilitated (Cheng, Wu, Liao, & Chan, 2009; ter Vrugte, de Jong, Vandercruysse, Wouters, van Oostendorp, & Elen, 2015). Last, challenge and complexity elements may have further impacted the flow experience (Inal & Cagiltay, 2007).

While most GBL studies examining the cognitive benefits of GBL have focused on learning outcomes at the completion of gameplay (e.g., Hwang, Wu, & Chen, 2012), they have largely ignored learning behaviors, except for a limited number of studies (e.g., Admiraal et al., 2011; Nietfeld, Shores, & Hoffmann, 2014). With the advancement of technology, learning behaviors can often be tracked and analyzed to provide valuable insight into GBL processes that lead to eventual learning outcomes (Cheng, Lin, & She, 2015). Accordingly, one purpose of the current study was to take into consideration learners’ learning behaviors while playing a science problem-solving game. Therefore, learners’ learning behavioral patterns in both conditions were analyzed and compared. We found that both conditions were similar in the way students frequently used means-ends and guessing strategies and actively sought help from the NPC. The differences were students in the non-competition condition tended to not only interact with the NPC, but to repeatedly seek additional support. These students also preferred to read instructions before making observations in the game, whereas students in the competition condition mainly relied on surface learning that included trying different variables and observing the effects; furthermore, they seldom sought help because they would have had to spend additional time reading and
following the instructional supports. Basically, student behavioral patterns in the competition condition were more focused on necessary movement, and they were less likely to explore other functions provided within the game environment. The reason for these behavioral differences may be due to the design of the competition and the various types of learners. In the competition condition, the students’ attentions were drawn to the in-game performance and the comparison of scores with their peers; the scoreboard showed the ranking and time spent on completing the game. Such competition may make students focus their attention on the scores and make them less engaged in the additional learning materials provided in the game. In this study, adding the ranking or competing scores may have subverted learning improvement. Additionally, differences in the student’s self-efficacy beliefs may have influenced their game learning behavior. Previous research stated that failure in a competition debilitates the greater efforts in learning for students with low self-efficacy beliefs (Bandura & Locke, 2003). These findings add to the existing literature that has explained how competition may harm the learning process by turning a project into a race to the finish line where understanding and internalizing concepts and knowledge becomes unimportant compared to winning. In GBSL, learners move from engagement in the form of making sense of the elements of the process to attempting to interpret and make a quality effort to efficiency, speed, and the outcome relative to others. The emphasis on the mode of competition, where a player sees his or her game performance as a contest, leads to a helpless pattern. This may further bring negative effects on student’s self-efficacy beliefs, motivation, and performance (Bandura & Locke, 2003; Van Eck & Dempsey, 2002). On the other hand, students in the non-competition condition played the game without any time constraints or peer pressure. Consequently, viewing a score without any comparisons may increase game involvement and may strengthen beliefs that putting effort in the game is worthwhile.

Conclusions and implications for future study

The role of competition in GBL has received increasing attention. However, the literature has overlooked how the level of competition in GBL grows abruptly. The results of this study provided new insights into the complex relationship between the inclusion of competition in a GBL environment and its effects on student learning achievement, flow experience, and learning behavioral patterns. Students in the non-competition condition showed significantly better learning achievement than those in the competition condition, while no significant difference was found in their flow experience. The results of the learning behavioral analyses revealed that, while both conditions resulted in students’ utilizing means-ends strategies, students in the non-competition condition tended to read the instructions carefully and repeatedly sought additional support to help themselves advance their conceptual understanding.

The results of this study offer several implications for game researchers, practitioners and designers. First, while competition condition of this study did not have significant positive effects, we recommend further investigation into other types of competition. For example, Chen, Law, and Chen (2018) suggested students might benefit more from the inclusion of anonymous competition. In our opinion, this could better promote learning, induce positive motivational outcomes, and facilitate meaningful cognitive engagement. Second, other possible effects of competition can be further examined, for example, the effects on interest, self-efficacy beliefs, or frustration. Third, the design of the game can explore the inclusion of intergroup competition on student learning achievement or learning behavioral patterns. For example, team-based competitive approaches may be especially effective in making instructional materials more enjoyable and engaging. Competition between groups may increase cooperation within groups because students are unified in working toward a common goal.

Acknowledgements

This research was supported by the Ministry of Science and Technology in Taiwan through the contract number MOST 105-2628-S-018-001-MY3. The authors would like to acknowledge the contributions of Yu-Kai Zhang, Migema Design Co., Ltd, and Jhih-Ping Jiang who supported the game development and research study.

References


A Votable Concept Mapping Approach to Promoting Students’ Attentional Behavior: An Analysis of Sequential Behavioral Patterns and Brainwave Data

Jerry Chih-Yuan Sun1*, Gwo-Jen Hwang2, Yu-Yan Lin1, Shih-Jou Yu1, Liu-Cheng Pan1 and Ariel Yu-Zhen Chen1

1Institute of Education, National Chiao Tung University, Taiwan // 2Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taiwan // csun@ntu.edu.tw // gjhwang.academic@gmail.com // aoisora.nagi@gmail.com // logotowa@gmail.com // bulaike1234@gmail.com // Ariel.Chen1216@gmail.com

*Corresponding author

ABSTRACT
This study explores the effects of integrated concept maps and classroom polling systems on students’ learning performance, attentional behavior, and brainwaves associated with attention. Twenty-nine students from an Educational Research Methodology course were recruited as participants. For data collection, in-class quizzes, attentional behavior analysis, and a 20-minute structured interview were applied, and the attention-associated brainwaves of the participants were measured. In the first week, a group-polling method was introduced in class; in the second week, participants were asked to draw concept maps using pen and paper (PnP concept mapping); and in the third week, the polling system and concept maps were integrated (votable concept mapping) and applied. The results showed that the PnP concept mapping approach improved the quiz results of students with lower learning motivation prior to the course, while the votable concept mapping method was effective in stimulating students’ attention during class. It was therefore suggested that instructors adopt methods integrating concept maps and polling tools to stimulate students’ attention and thereby promote a positive cycle of attentional behavior in the classroom. For example, students’ attentional behavior during an activity facilitated their attentional behavior after the activity, and this behavior continued until the next activity.

Keywords
Concept map, Polling, Attentional behavior, Attention, Brainwave

Introduction

The purpose of this study is to explore the effects of the integration of concept maps and classroom polling systems on students’ learning performance, brainwaves associated with attention, and behavioral sequences. Concept mapping has been applied to a variety of educational settings, such as classroom teaching (Chiou, Lee, & Liu, 2012; Sun & Lee, 2016), designing digital teaching materials (Adesope & Nesbit, 2013), in-class teaching activities (Jones, Ruff, Snyder, Petrich, & Koonce, 2012), in-field inquiry activities (Hwang, Wu, & Ke, 2011), and online inquiry activities (Hwang, Kuo, Chen, & Ho, 2014). Researchers have indicated that concept maps can facilitate learners’ reasoning ability (Mih & Mih, 2011); moreover, providing well-constructed concept maps during the learning process improves the accuracy of learners’ understanding of the knowledge (Redford, Thiede, Wiley, & Griffin, 2012). On the other hand, researchers have further emphasized the importance of facilitating peer interactions (Sun, Chen, Yeh, Cheng, & Lin, 2018), which are usually ignored in most concept mapping activities. Using a polling tool in the classroom helps teachers to attract students’ attention, increase their engagement level, and obtain information regarding their understanding of the subject. In addition, the application of the polling system facilitates peer interaction; learners are able to compare their answers with those of their classmates, discuss discrepancies, and reassess their own answers (Gachago, Morris, & Simon, 2011). Therefore, this study combines the concept-mapping technique and a polling system installed on tablets to help learners to strengthen the links between concept nodes through the introduction of interactive voting activities into the construction processes of concept maps.

It is generally considered that bringing sufficient attention to learning activities ensures meaningful learning. In the study of Hwang, Yang, and Wang (2013), researchers introduced the concept mapping technique to a game-based learning system and found that the teaching method resulted in learners becoming better focused and taking greater initiative in learning activities. Nesbit, Larios, and Adesope (2007) utilized eye movement tracking devices to examine the manner in which students read concept maps. The results proved that well-designed concept maps optimize the efficiency of attention allocation.

Sequential behavioral analysis is a behavior research method that utilizes encoded behaviors to investigate time-based behavioral patterns of individuals and groups (Bakeman & Gottman, 1997, p. 14). Hou (2012a) compiled a
log of students’ operations on a large-scale multi-person online educational gaming platform to analyze learners’ knowledge construction, peer interaction, and problem-solving processes. Hou (2015) investigated learners’ behavioral patterns and flow states in game-based learning to understand the patterns of their interactive behavior during the learning process. On that account, the sequential behavioral analysis technique could be used to explore students’ interactive behavior in the class, thereby leading to an in-depth understanding of learners’ knowledge construction processes. Therefore, this study adopted the sequential behavioral analysis method to explore learners’ attentional behavior during the concept mapping process.

To summarize, this study introduced a classroom polling system to the construction of concept maps in order to stimulate peer interaction, reflection, and discussion, improve students’ understanding of the course knowledge, and enhance their attention during learning. The research framework of the present study is shown in Figure 1. The research questions are as follows:

- Are there any significant differences in the academic performance of learners with different motivational traits when different conceptual mapping strategies and tools are applied?
- Are there any significant differences in the patterns of learners’ attentional behaviors when different conceptual mapping strategies and tools are applied?
- Are there any significant differences in the brainwave readings associated with attention when different conceptual mapping strategies and tools are applied?

![Figure 1. Research framework](image)

**Literature review**

**Integrating conceptual maps and the polling system, and the role of learning motivation**

A concept map is a graphical tool that illustrates the relationships between concepts. Concept maps use nodes (usually circles and squares) to represent concepts, links (connecting lines) to indicate the relations between nodes, and labels (texts and symbols) to describe their relationships (Novak, 1984). Through systematic induction and organization, concept maps transform large-scale, complex knowledge into a visual map to help learners better understand the meaning of each concept through the layers and links between the nodes (Blankenship & Dansereau, 2000; Novak, 1990).

Polling systems (also known as “clickers”) refer to a small voting tool that is commonly used during teaching. Specifically, teachers first use presentation tools (such as MS PowerPoint) to introduce a question and corresponding options to the students, and then use a polling system to collect anonymous votes on each option on behalf of the students. Polling systems are used to analyze learners’ understanding of the knowledge given; they also serve as an in-class quiz and assessment tool and record learners’ attendance (Cheesman, Winograd, & Wehrman, 2010; Prather & Brissenden, 2009; Sun, Martinez, & Seli, 2014). Gachago et al. (2011) pointed out that the application of classroom polling systems is conducive to attracting learners’ attention and enhances their engagement.
Many scholars have combined concept mapping and technological devices in their research and found that the combination had a positive effect on learning results. The study of Hwang et al. (2014) showed that a fill-in-the-blank concept map improved learners’ problem-solving abilities. Hwang et al. (2011) applied concept mapping to an outdoor learning activity by presenting the teaching material with concept maps on mobile devices, which allowed the learners to draw and modify the concept maps. The results showed that the teaching method enhanced learners’ comprehension of the knowledge. However, if learners were not familiar with concept maps, they tended to have difficulties understanding the structural relations between concepts, which thereby affected their learning motivation (Blankenship & Dansereau, 2000). Sun and Lee (2016) found that, compared with students with low learning motivation prior to the course, students with high learning motivation tended to develop greater motivation for learning after participating in a course that used tablet computers to construct concept maps. For that reason, participants’ learning motivation before the in-class quiz was measured, and cluster analysis was applied to divide the participants into groups; a comparative analysis was then employed to investigate the differences in the learning achievements of the different groups.

Expectancy-value theory is one of the most important theories related to learning motivation. Expectancy refers to individuals’ beliefs and judgments regarding their ability to successfully complete a given task. Value refers to the incentive that drives individuals to engage in the task (Schunk, Meece, & Pintrich, 2013). This study used participants’ performance expectations of the in-class quiz as a proxy variable for their motivational traits. In addition, since we only needed to measure participants’ performance expectations in the quiz, a single-item questionnaire based on the definition of expectancy proposed by Schunk et al. (2013) was adopted.

One other notable learning motivation related factor is anxiety. Evaluative environments and timed test-taking conditions (such as quizzes or exams) accentuate the detrimental effects of anxiety. The study of Tsai, Lin, and Yuan (2001) revealed that learners with greater test anxiety are more likely to prefer to use the developed fill-in-the-blank concept maps. Batchelor (2015) introduced clickers in a calculus course and discovered that learners’ engagement in the learning process and their expectations of the examination results are powerful predictors of their math anxiety. According to the aforementioned studies, anxiety is an indicator of affectivity and is associated with performance expectation. Therefore, anxiety was included as a variable in the cluster analysis of motivational traits. Considering the context of the in-class quiz in this study, a single-item questionnaire based on the definition of test anxiety proposed by Zeidner (1998) was used to measure participants’ test anxiety.

### Attentional behavior and attention-associated brainwaves during a votable concept mapping activity

Engaging sufficient attention in learning tasks is essential to achieving meaningful learning. Existing empirical studies have revealed that concept mapping techniques and classroom polling tools have a positive effect on students’ engagement and attention in learning. Hwang et al. (2013) found that combining concept maps and game-based teaching tends to make learners more attentive in class and to participate more actively in the learning activities. The study of Nesbit et al. (2007) applied eye-movement tracking devices and discovered that a well-designed concept map can improve the efficiency of attention, allowing learners to allocate attentional information resources from the higher part of the hierarchy and the center of the network to more effectively complete knowledge construction. Sun (2014) employed physiological equipment to trace learners’ brainwaves, and found that their attention was significantly enhanced during the voting process. Therefore, this study designed votable concept maps to enhance learners’ attentional behavior and attention-related brainwaves through the interaction of the voting and knowledge construction processes of concept mapping.

Physiological data, such as detecting electrical activity in the brain with electroencephalographic (EEG) devices (Rebolledo-Mendez et al., 2009), reflects learners’ attention levels from an objective perspective. Referring to the review of Mcdowd (2007), this study attempts to measure learners’ attention from both a behavioral and physiological approach so as to acquire an in-depth understanding of changes in their attention level. Specifically, all participants’ behaviors during the class were recorded for further encoding, and the brainwaves of three participants were measured throughout the entire experiment process.

Sequential behavioral analysis is a commonly applied method to investigate learning behavior that reveals the behavioral patterns of individuals and groups based on the sequence of encoded behaviors (Bakeman & Gottman, 1997; Sun, Kuo, Hou, & Lin, 2017). The process of learning how to use concept maps is considered a knowledge construction behavior, and can be analyzed by sequential behavioral analysis to gain a better understanding of how concept mapping enhances learning. Hwang et al. (2011) investigated the effect of mobile-facilitated concept mapping strategies on students’ learning achievements and attitudes in ecology courses. Hwang et al. (2014) developed a computer-supported concept map-based teaching system which improves
learners’ problem-solving ability. Sun and Chen (2016) incorporated dynamic concept maps into a polling system, and investigated its effect on students’ learning motivation and achievement. However, the majority of the studies that combine concept maps and technological devices focus on the results of learning and learning motivation, so the actual sequence of students’ attentional behaviors during the concept mapping processes has yet to be explored. Therefore, this study uses sequential behavioral analysis to investigate participants’ attentional behaviors during the voting and concept mapping activities.

In summary, previous studies have revealed that both concept mapping and polling systems have positive effects on learning performance, while the integration of concept maps and technological devices can further improve learners’ comprehension of the knowledge (Hwang et al., 2014; Hwang et al., 2011). For that reason, this study hypothesized that polling systems can stimulate interactions in learning so that learners are able to construct more complete concept maps and thereby improve the learning effectiveness. The study of Sun and Lee (2016) revealed that learners’ motivational traits influence the effectiveness of teaching strategies for tablet-facilitated concept mapping. Therefore, we also included performance expectations and anxiety as proxy variables for learning motivation in our study. Furthermore, since polling systems stimulate interactions in learning, learners’ attention during the concept map construction process could be enhanced, which may lead to increased attentional behavior and enhanced attention-related brainwaves.

Research methods

Participants

The participants of the study were 33 graduate students registered in an educational research methodology course. Their in-class quiz results, attentional behavior, and attention-related brainwaves during the research period were collected. Excluding the data of students who did not complete the three-week session or provide the required information, the data of 29 participants were retained for analysis. In addition, the brainwave readings of three participants were collected for data analysis to gain a deeper understanding of the brainwave variations over the 3 weeks. The three participants (two males and one female) were volunteers who had given their consent to the collection of their brainwave data. The participants were divided into groups of three to four during the voting activities. Among the 29 participants, nine were male (31%) and 20 were female (69%). The average age was 26.34 years old ($SD = 5.88$).

Instructional design

The study lasted 3 weeks; each week involved 100 minutes of classes. Each class included a lecture and two to three group activities. Three participants were selected and their brainwaves monitored throughout the entire 3-week period. Twenty minutes before the end of the last class of each week, participants were asked to rate their feelings of anxiety and performance expectations regarding the upcoming quiz; an in-class closed-book quiz was then conducted. In addition, at the end of the 3-week period, nine participants were invited to participate individually in a 20-minute face-to-face structured interview.

Three group activities were implemented during the research period. Specifically, the voting activities and use of tablets were implemented in week 1; the pen-and-paper (PnP) concept mapping activities were adopted in week 2; and votable concept maps were introduced in week 3. The polling system application employed by the study was the “Interactive, Feedback-Based In-Class Teaching” (iFIT) system, which has been applied to classroom voting activities and achieved positive results prior to our study (Sun et al., 2018; Sun & Lee, 2016). In the voting activity, a tablet was provided to each group. After the instructor finished a section of the teaching material, he would present a question related to the knowledge taught on the screen for group discussion. The participants were asked to discuss the options and use the provided tablet and application to vote anonymously for the answer that they most preferred. The votes were then transmitted to the back of the instructor’s stage area, the total for each answer calculated automatically by the polling system, and the results presented on the screen. The instructor would then reveal the correct answer, explain the reasons, and help students to resolve any misunderstandings or confusion.
Figure 2. Polling questions used for group discussion

Figure 3. Experimental design
In the PnP concept mapping activity, the instructor would also present a question on the screen after each section. The participants were required to exchange viewpoints and ideas with their group and construct a concept map using the pen and paper provided to each group. Next, each group would take a photo of the constructed concept map with the tablet provided and send the photo to the instructor. The instructor would then display the concept map of each group on the screen, discuss the strengths and weaknesses of each map, and resolve any misunderstandings and errors the class might have. In the votable concept mapping activity, after each section, the instructor would present the question in the format of a fill-in-the-blank concept map with
corresponding options. The participants were required to choose the best answer for each blank following a group discussion. The voting results would be presented to the class and the instructor would reveal the correct answer, explain the reasons, and resolve any misunderstandings or confusion related to the concepts. The questions used in the group discussion over the 3 weeks are shown in Figure 2. In order to avoid the carry-over effect, the contents of the three units (“threats to validity,” “experimental design,” and “research method”) were independent, with no sequenced relationship between two units. The difficulty level of the quizzes was approximately equal. The procedure of the experiment is presented in Figure 3 and photos taken during the experiment are shown in Figure 4.

**Instruments and Analysis**

The item used to measure performance expectations was phrased as follows: “In the upcoming quiz, I expect my score to be ___ (from 0 to 100).” The item used to measure anxiety was written as follows: “From 0 to 100, what is your anxiety level regarding the upcoming quiz?” The in-class quizzes were provided by the instructor and included multiple-choice, fill-in-the-blank, and short-answer questions. Due to differences in the rated scores, the Z-scores of performance expectations and anxiety of the participants were computed, and cluster analysis was then applied to divide the participants into groups. In the first stage of the cluster analysis, Ward’s method was used to determine the optimal number of clusters. In the second stage, the k-means clustering technique was applied to determine the allocation of participants among the groups, and yielded three groups: the “high-expectancy and high-anxiety group” (Group 1), the “medium-expectancy and low-anxiety group” (Group 2), and the “low-expectancy and medium-anxiety group” (Group 3); the number of participants in each group was 7, 9, and 13, respectively. Next, we applied Analysis of Variance (ANOVA) to test the validity of the clustering results. The results showed significant differences in both the performance expectations ($F = 18.11, p < .001$) and anxiety ($F = 20.26, p < .001$) of the three groups. The means of performance expectations and anxiety of Group 1 were 81.86 ($SD = 9.70$) and 85.71 ($SD = 11.34$), respectively; those of Group 2 were 74.78 ($SD = 8.97$) and 42.22 ($SD = 17.87$), respectively; and those of Group 3 were 57.69 ($SD = 9.27$) and 71.92 ($SD = 12.84$), respectively. The findings revealed distinctive differentiations in the motivational traits of the three groups, indicating that the clustering results had satisfactory validity. Due to the small size of the research sample, we adopted the Kruskal-Wallis H test, a nonparametric statistical method, instead of ANOVA to examine the significance level of the dissimilarities in the quiz results of the three groups.

<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Behaviors that show attention to the instructor</td>
<td>Looking at the instructor, listening while the instructor is talking</td>
</tr>
<tr>
<td>2</td>
<td>Behaviors that show attention to the teaching materials</td>
<td>Checking the teaching material, taking notes</td>
</tr>
<tr>
<td>3</td>
<td>Behaviors involving interaction with the tablet</td>
<td>Using the tablet to answer questions and research information</td>
</tr>
<tr>
<td>4</td>
<td>Behaviors involving interaction with group members</td>
<td>Talking and discussing with group members</td>
</tr>
<tr>
<td>5</td>
<td>Behaviors that show attention to the concept map</td>
<td>Constructing and modifying the concept map on paper (referring to the activity in the week of PnP concept map)</td>
</tr>
<tr>
<td>6</td>
<td>Behaviors that show attention to the instructor’s explanation of the answers</td>
<td>Listening to the instructor’s explanation, checking the textbook, checking the teaching materials, looking at the presentation</td>
</tr>
<tr>
<td>7</td>
<td>Distracted behaviors involving digital devices</td>
<td>Using any irrelevant digital devices in class (such as smart phones and MP3s)</td>
</tr>
<tr>
<td>8</td>
<td>Distracted behaviors involving classmates</td>
<td>Chatting with classmates on irrelevant topics</td>
</tr>
<tr>
<td>9</td>
<td>Other distracted behaviors</td>
<td>Looking around the room (not at the textbook and teaching material), staring blankly into space, searching for irrelevant items</td>
</tr>
</tbody>
</table>

We outlined the coding scheme of the attentional behavior of the participants. According to the learning tasks and required attention in the experiment, the behaviors were coded into three categories: “attentional behavior
related to the course,” “attentional behavior related to the activity,” and “distracted behavior.” The detailed coding method of the behaviors is shown in Table 1. Three observers were in charge of the behavior coding process by observing participants’ behaviors recorded during the research period. The coded behaviors of three students during the polling system week were used to examine the reliability of the encoding process. Fleiss’ kappa is a statistical indicator of the reliability of agreement between three raters when handling nominal scales (Fleiss, 1971). The Fleiss’ kappa of the encoding process of the study was 0.65, suggesting that the consistency between the raters was substantial (Landis & Koch, 1977, p. 165).

Brainwave signals of three students were measured to monitor their attention level during the classes. NeuroSky MindWave headsets were used in this study, as they have been proven able to effectively monitor users’ attention levels through brainwave signals (Crowley, Sliney, Pitt, & Murphy, 2010; Rebolledo-Mendez et al., 2009). Based on the nature of the teaching activities, we divided the entire course into three stages, classroom lectures, group activities, and explaining the answers after the activity, and analyzed the EEG data at each stage accordingly. We used the first 3 minutes of the EEG data of each participant as the baseline for their attention levels, then calculated the percentage of the attention levels that exceeded the baseline and plotted the EEG diagrams for the selected participants accordingly. Lastly, we combined the EEG observations with the results of the sequential analysis of attentional behavior to further explain the variations in brainwaves.

Results

Comparison of the quiz results of the three groups

The Kruskal-Wallis test results showed no significant differences in the quiz results of the three groups during the polling system week ($\chi^2 = 3.09, p = .21$). However, significant differences were found in the quiz results of the three groups in the PnP concept map week ($\chi^2 = 6.56, p = .04$). A post-hoc comparison revealed that the mean rank of quiz scores of Group 3 (12.46) was significantly better than that of Group 1 (12.21). The quiz results of the three groups in the votable concept map week showed no significant differences ($\chi^2 = 1.23, p = .54$).

Sequence of attention behavior over the 3 weeks

The adjusted residuals (Z-scores) of the originally coded behavioral data are presented in Table 2. In the table, the initial behaviors are listed in the second column of each row, and subsequent behaviors in the first row of each column. If the Z-score between two behaviors was greater than 1.96, then the sequential relationship between the behaviors can be considered statistically significant ($p < .05$) (Bakeman & Gottman, 1997; Hou, 2012b). The sequence patterns of participants’ attentional behaviors during the 3 weeks are demonstrated in Figure 5.

As shown in the figure, in the polling system week, participants were found to be engaged in group discussion after the voting activities, and the behavioral sequence of “discussing following voting” (3 → 4) reached statistical significance. However, the behavioral sequence of “voting following discussion” (4 → 3) was not statistically significant. Since during the PnP concept map week the tablets were used to take and send photos rather than being used as voting instruments, no significant sequential relations were found between interaction with group members (4) and interaction with the tablet (3). Instead, attention to the concept map (5) and interaction with group members (4) were found to have significant sequential relations in both directions (4 → 5 and 5 → 4). In addition, the participants were found to show attention to the instructors’ explanation of the answer and analysis of the concept map constructed by each group (3 → 6), and the sequential relationship was statistically significant. However, participants were also found to be shifting between listening to the instructor and other distracted behaviors (6 → 9, 9 → 6), with notable distracted behaviors, such as looking around the room, staring blankly into space, and looking for irrelevant items. The behavioral sequence of “discussing the following votes” (3 → 4) was also found to be statistically significant in the votable concept map week. Moreover, interaction with the tablet (3) and interaction with group members (4) were both found to lead to attention paid to the instructor’s explanation of the answers (3 → 6 and 4 → 6), while listening to the instructor’s explanation of the answers (6) was also found to effectively enhance the interaction with the tablet and group members in the next activity (6 → 3 and 6 → 4). The attentional behaviors during the activity and following the activity in the votable concept map week appeared to be able to facilitate one another, thus forming a virtuous cycle.
Table 2. Z-Score table of participants’ attentional behavior during the 3 weeks

<table>
<thead>
<tr>
<th>Weeks</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>
</tr>
</thead>
<tbody>
<tr>
<td>Polling System</td>
<td>1</td>
<td>115.77</td>
<td>-76.14</td>
<td>-25.12</td>
<td>-33.26</td>
<td>This</td>
<td>-44.75</td>
<td>-6.44</td>
<td>-19.63</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-15.95</td>
<td>-14.16</td>
<td>115.55</td>
<td>3.34*</td>
<td>only</td>
<td>-2.86</td>
<td>-2.28</td>
<td>-3.86</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-21.00</td>
<td>-18.07</td>
<td>0.95</td>
<td>116.15</td>
<td>appears</td>
<td>-1.37</td>
<td>-3.02</td>
<td>-5.92</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>-25.61</td>
<td>-22.37</td>
<td>-6.17</td>
<td>-8.27</td>
<td>in the</td>
<td>126.99</td>
<td>-3.82</td>
<td>-7.03</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>-3.81</td>
<td>-6.61</td>
<td>-1.78</td>
<td>-2.53</td>
<td>PnP</td>
<td>-3.26</td>
<td>95.55</td>
<td>-1.77</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>-11.48</td>
<td>-10.20</td>
<td>-3.48</td>
<td>-5.86</td>
<td>concept</td>
<td>-7.02</td>
<td>-2.24</td>
<td>85.94</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>-12.46</td>
<td>-11.95</td>
<td>-4.02</td>
<td>-3.10</td>
<td>map week</td>
<td>-6.56</td>
<td>-0.80</td>
<td>-1.22</td>
</tr>
<tr>
<td>PnP Concept Map</td>
<td>1</td>
<td>95.15</td>
<td>-43.20</td>
<td>-8.11</td>
<td>-40.36</td>
<td>-17.23</td>
<td>-35.73</td>
<td>-8.67</td>
<td>-9.14</td>
</tr>
<tr>
<td>Concept Map</td>
<td>2</td>
<td>-44.45</td>
<td>100.87</td>
<td>-8.80</td>
<td>-43.75</td>
<td>-18.41</td>
<td>-38.14</td>
<td>-10.56</td>
<td>-13.30</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-5.67</td>
<td>-5.07</td>
<td>98.13</td>
<td>-0.53</td>
<td>-1.34</td>
<td>3.01*</td>
<td>-0.78</td>
<td>-1.73</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-30.21</td>
<td>-31.45</td>
<td>1.80</td>
<td>124.78</td>
<td>7.76*</td>
<td>-4.88</td>
<td>-4.03</td>
<td>-8.85</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>-11.46</td>
<td>-11.61</td>
<td>0.17</td>
<td>5.60*</td>
<td>112.69</td>
<td>-4.78</td>
<td>-1.49</td>
<td>-3.34</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>-24.96</td>
<td>-25.12</td>
<td>0.12</td>
<td>-6.64</td>
<td>-5.69</td>
<td>122.77</td>
<td>-3.30</td>
<td>-6.89</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>-5.55</td>
<td>-6.56</td>
<td>-0.78</td>
<td>-3.63</td>
<td>-1.48</td>
<td>-3.07</td>
<td>136.20</td>
<td>-1.91</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>-4.68</td>
<td>-9.29</td>
<td>-1.78</td>
<td>-8.04</td>
<td>-3.38</td>
<td>-6.72</td>
<td>-1.96</td>
<td>80.57</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>-7.11</td>
<td>-13.09</td>
<td>-2.31</td>
<td>-5.77</td>
<td>-4.39</td>
<td>2.34*</td>
<td>-2.11</td>
<td>1.50</td>
</tr>
</tbody>
</table>
| Table 2. Z-Score table of participants’ attentional behavior during the 3 weeks

Note. 1 = Behaviors that show attention to the instructor; 2 = behaviors that show attention to the teaching materials; 3 = behaviors involving interaction with the tablet; 4 = behaviors involving interaction with group members; 5 = behaviors that show attention to the concept map; 6 = behaviors that show attention to the instructor’s explanation of the answers; 7 = distracted behaviors involving digital devices; 8 = distracted behaviors involving classmates; and 9 = other distracted behaviors. *p < .05.

![Figure 5. Sequence patterns of participants’ attentional behavior during the 3 weeks](image)

Note. 1 = Behaviors that show attention to the instructor; 2 = behaviors that show attention to the teaching materials; 3 = behaviors involving interaction with the tablet; 4 = behaviors involving interaction with group members; 5 = behaviors that show attention to the concept map; 6 = behaviors that show attention to the instructor’s explanation of the answers; 7 = distracted behaviors involving digital devices; 8 = distracted behaviors involving classmates; and 9 = other distracted behaviors. *p < .05.
members; 5 = behaviors that show attention to the concept map; 6 = behaviors that show attention to the instructor’s explanation of the answers; 7 = distracted behaviors involving digital devices; 8 = distracted behaviors involving classmates; and 9 = other distracted behaviors.

**Differences in the brainwave signals over the 3 weeks**

Table 3 and Figures 6 and 7 illustrate the percentage of the brainwave signals collected from the three participants (A, B, and C) that exceeded the baseline (hereinafter, attention indicator). Figure 6 is a comparison of the week-on-week changes in the attention indicator. It can be seen from the figure that A’s and C’s attention indicators during the classroom lectures were the highest in the PnP concept map week (50% greater than the baseline), while B’s attention indicator was the highest in the polling system week (60% greater than the baseline). In the group activities, the attention indicators of A, B, and C were the highest in the PnP concept map week, the votable concept map week, and the polling system week, respectively. When the instructor was explaining the answers following the activities, the attention indicators of A and B were highest in the PnP concept map week, while that of C was highest in the polling system week. From the perspective of the entire class, the attention indicators of A and C were highest in the PnP concept map week and that of B was highest in the polling system week. Figure 7 compares the three participants’ attention indicators at different stages of the course in the same week. In the polling system week, the attention indicators of A and B were highest in classroom lectures (30% and 50% greater than the baseline, respectively), while that of C was highest when the instructor was explaining the answers after the activities (70% greater than the baseline). In the PnP concept map week, A’s attention indicator was highest during the classroom lectures, B’s during the group activities, and C’s when the instructor explained the answers after the activities. In the votable concept map week, both A’s and B’s attention indicators were highest during the group activities, while C’s remained highest when the instructor explained the answers after the activities.

**Table 3. Frequency of participant attention greater than the baseline**

<table>
<thead>
<tr>
<th>Participants</th>
<th>Weeks</th>
<th>Stages of the course</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Classroom Lecture</td>
<td>Group Activity</td>
<td>Explaining the Answers after the Activity</td>
<td>The Whole Course</td>
</tr>
<tr>
<td>A</td>
<td>Polling System</td>
<td>34.43%</td>
<td>26.27%</td>
<td>10.32%</td>
<td>31.64%</td>
</tr>
<tr>
<td></td>
<td>PnP Concept Map</td>
<td>73.46%</td>
<td>49.83%</td>
<td>41.18%</td>
<td>60.63%</td>
</tr>
<tr>
<td></td>
<td>Votable Concept Map</td>
<td>16.56%</td>
<td>31.29%</td>
<td>26.32%</td>
<td>20.90%</td>
</tr>
<tr>
<td>B</td>
<td>Polling System</td>
<td>52.90%</td>
<td>46.24%</td>
<td>36.52%</td>
<td>49.26%</td>
</tr>
<tr>
<td></td>
<td>PnP Concept Map</td>
<td>33.20%</td>
<td>48.40%</td>
<td>38.46%</td>
<td>35.88%</td>
</tr>
<tr>
<td></td>
<td>Votable Concept Map</td>
<td>35.41%</td>
<td>54.26%</td>
<td>29.41%</td>
<td>36.60%</td>
</tr>
<tr>
<td>C</td>
<td>Polling System</td>
<td>53.03%</td>
<td>61.83%</td>
<td>73.33%</td>
<td>57.62%</td>
</tr>
<tr>
<td></td>
<td>PnP Concept Map</td>
<td>63.31%</td>
<td>41.12%</td>
<td>70.08%</td>
<td>62.62%</td>
</tr>
<tr>
<td></td>
<td>Votable Concept Map</td>
<td>46.70%</td>
<td>40.36%</td>
<td>61.39%</td>
<td>46.72%</td>
</tr>
</tbody>
</table>

*Figure 6. Week-by-week comparison of three participants’ attention indicators during each course stage*

Note. The stars highlight the week when the participants’ attention indicators were highest in a given stage.
Figure 7. Comparison of three participants’ attention indicators during all course stages in each week
Note. The stars highlight the stage when the participants’ attention indicators were highest in a given week.

Figure 8. Changes in participant B’s attention levels during teaching activities in the polling system and votable concept map weeks

Figure 8 illustrates the changes in B’s attention indicator during the teaching activities in the polling system and votable concept map weeks. It can be seen from the figure that in the polling system week, this indicator was unstable during the group activities as well as when the instructor was explaining the answers after the activities. In the votable concept map week, fewer fluctuations in the indicator were observed during the group activities and when the instructor was explaining the answers after the activities, indicating that B’s attention level was relatively more stable in the votable concept map week than in the polling system week.
In general, the participants’ attention level was highest in the PnP concept map week. Participants were found to invest more attention in the lecture and the explanations after the group activity, when a PnP concept mapping activity was implemented. In weekly terms, the participants’ attention levels were highest during the classroom lecture stage in the polling system week and the group activity stage in the votable concept map week; however, no specific stage received additional attention from the participants in the PnP concept map week. In addition, compared with the voting activities in the polling system week, participants showed a more stable level of attention during the votable concept mapping activities, as well as when the instructor was explaining the answers after the activities in the votable concept map week (Figure 8).

Discussion

The purpose of this study was to integrate the concept mapping technique and polling system as a teaching strategy and explore their impact on in-class quiz results, attentional behavior, and brainwaves associated with the attention of students with various motivational traits. The results showed that the mean rank of the quiz results of the low-expectancy and medium anxiety group (12.46) was significantly better than that of the high-expectancy and high-anxiety group (12.21). Past research has confirmed that combining concept mapping techniques and technological devices that provide learners with the opportunity to construct concept maps in a fill-in-the-blank format can improve learners’ problem-solving ability and enhance their comprehension of the knowledge (Hwang et al., 2014; Hwang et al., 2011). However, the study of Sun and Lee (2016) suggested that, compared with concept mapping activities that use tablet computers, concept mapping activities that use pen and paper are able to significantly improve the post-test motivation of learners with lower initial motivation. Similar results were obtained in this study: The PnP concept mapping activity was found to be able to improve the quiz scores of participants from the low-expectancy and medium anxiety group. Nesbit et al. (2007) found that concept maps could help learners to improve the utility of attentional resources and construction of knowledge. Therefore, this study speculated that the results could be attributed to the scaffolding structure of concept maps, which can help learners with low self-confidence and a certain level of anxiety allocate their attentional resources more efficiently. Thus, they are able to participate in the group discussions more actively and construct concept maps with other group members. As a result, their learning performance was improved.

The results of this study revealed that during the first week, when the voting activities were introduced, there was a significant sequential relation between voting and group discussion (“discussing after voting”), and the brainwave readings showed that classroom lectures received the highest level of attention. During the second week, when the PnP concept mapping activities were implemented, an apparent repetition of two behaviors, “discussing with group members” and “constructing concept maps,” was observed, showing that the concept maps were completed through multiple discussions among group members. In addition, after taking photographs of the constructed concept maps and sending them to the instructor, participants were found to pay greater attention to the instructor’s explanation of the answers. However, while listening to the instructor’s explanation, participants also manifested “other distracted behaviors.” The brainwave readings showed that classroom lectures received the highest level of attention. During the third week, when votable concept mapping activities were applied, the behavioral sequence of “discussing after voting” was observed again. In addition, interactions with both the tablet and group members were found to lead to greater attention shown to the instructor’s explanation after the activities. More importantly, it was observed that listening to the instructor’s explanation tended to lead participants to interact with the tablet and group members in the next activity. These findings show that votable concept maps can stimulate learners’ interactions and enhance their attention to the instructor’s explanations, forming a virtuous cycle of learning behavior. The brainwave readings showed that group activities received the highest level of attention throughout the week.

Sun (2014) discovered that learners’ attention-related brainwaves were significantly enhanced when participating in activities involving interactive response systems (IRSs). The same result was also reached in this study. To quote participant B’s comments on the votable concept maps during the interview, “...and then answer the question proposed by the instructor; for this stage, since there were images to supplement the information previously given by the instructor, [the concept of] this part became quite clear [to me].” It can be seen that a votable concept map can clearly present the knowledge structure and enhance learners’ understanding of the topics, and that participants’ attention-related brainwaves were significantly strengthened during the voting activity. One possible reason for this result is that the votable concept map can stimulate learners’ attention-related brain regions, which leads to their attentional behavior when listening to the instructor’s explanation and more active engagement in interactive discussions of the next activity. Hwang et al. (2013) discovered that combining concept maps and game-based teaching could effectively enhance learners’ learning achievement and mental effort in learning activities.
The findings in the PnP concept mapping activities of this study were the same. During the group activities, the learners were found to be switching between drawing concept maps and discussing with group members. However, it was also found that after the concept mapping activity had finished, learners manifested noticeable distracted behaviors, such as looking around the room, staring blankly into space, and looking for irrelevant items, when the instructor was explaining the answers. The results of the interviews revealed that participants were not satisfied with the clarity of the projected concept maps constructed by each group. For example, some participants claimed that “the instructor seemed to have presented ours on the screen, however, since the words are really small, we cannot see what they are about...”; “because the photos are not clear; [I] hope that other methods can be used [in the future]”; and “the concept maps drawn by us are not clear.” Thus, one likely explanation for the distracted behaviors after the group activity observed in the PnP concept map week is the lack of clarity of the concept maps projected on the screen, which led to the learners experiencing difficulties concentrating on the screen, thus leading to their distracted behaviors.

Conclusion and future research

This study investigated the effects of integrating concept maps and a polling system in teaching on learners’ quiz results, attentional behavior, and brainwaves associated with attention. The conclusions of the study are as follows: (1) When PnP concept mapping activities were implemented, the mean rank of the participants’ quiz scores from the low-expectancy and medium-anxiety groups were significantly better than that from the high-expectancy and high-anxiety group. (2) When PnP concept mapping activities were carried out in class, the learners were found to be highly attentive during the interactive activities; however, their attention was distracted when the instructor was explaining the answers. (3) During the class with the use of votable concept mapping activities, the attention level measured through brainwave signals was the greatest compared with all group activities. (4) During the class with the application of votable concept mapping activities, learners’ attentiveness to the discussions of one activity tended to lead to attentive behavior when the instructor was explaining the answers, which was likely to lead to their attentive behavior in the next discussion of the next activity, forming a virtuous behavioral cycle. In short, PnP concept mapping activities could improve the learning performance of students with low learning motivation, leading to attentional behavior and active participation in discussion and interaction. However, learners might be distracted when the instructor starts to explain the answers after the activities. Votable concept mapping activities, on the other hand, are not only conducive to promoting attentional behavior during learning activities, but also encourage learners to concentrate on the instructor’s explanation following the activities. On that account, votable concept mapping effectively enhances learners’ attentional behavior prior to and following the activities.

The limitations of this study include the small sample size (students from only one class were recruited as participants), the short research period, the limited number of participants used for brainwave data collection (only three students), and the limitations of the reliability and validity of the research instruments. Therefore, the findings of this study should be used with caution. We suggest that further studies expand the size of the research sample and extend the research period to improve the reliability of the results. In addition to introducing a control group in the research design, future studies can also attempt to use a counterbalanced design that randomly assigns participants into groups to minimize the carry-over effect and achieve more distinguishable results by implementing various votable concept-mapping activities in different orders. In future research, it is recommended that researchers include a preliminary test, so that the participants’ prior knowledge regarding each subject may be included in the analysis. Future studies involving votable concept maps are suggested in order to introduce more comprehensive measurement scales that can examine learners’ motivation and learning performance, so as to better understand the influence of the teaching method on the learners’ performance and different motivational traits. In terms of experimental devices, future studies should increase the number of wearable brainwave headsets so as to collect EEG data from more learners and conduct a more in-depth analysis of changes in their attention levels. Lastly, further studies can also incorporate group competition activities to stimulate learners’ participation and concentration and thereby enhance their attentional behavior in the learning process after the activities have been completed.

Acknowledgements

This research was supported by the Ministry of Science and Technology in Taiwan through Grant numbers MOST 105-2511-S-009-013-MY5 and MOST 104-2511-S-009-008-MY3. The authors would like to thank the students who participated in this study, Katherine Pin-Chen Yeh, Afra Lin, Hsueh-Er Tsai, Chih-Hsuan Chao, and
Yen-Ling Chen for their assistance in conducting this research, Shu-Kai Yang for his support for the smooth implementation of the study’s iFIT app, and all of the reviewers for their useful comments.

Reference


Using a Learner-Topic Model for Mining Learner Interests in Open Learning Environments

Pengfei Wu1,2,3, Shengquan Yu1,2* and Dan Wang1,2

1School of Educational Technology, Faculty of Education, Beijing Normal University, China // 2Advanced Innovation Center for Future Education, Beijing Normal University, China // 3School of Education, Shijiazhuang University, China // wupengfei_2000@163.com // yusq@bnu.edu.cn // woxinwing@163.com
*Corresponding author

ABSTRACT

The present study uses a text data mining approach to automatically discover learner interests in open learning environments. We propose a method to construct learner interests automatically from the combination of learner generated content and their dynamic interactions with other learning resources. We develop a learner-topic model to discover not only the learner’s knowledge interests (interest in generating content), but also the learner’s collection interests (interest in collecting content generated by others). Then we combine the extracted knowledge interests and collection interests to yield a set of interest words for each learner. Experiments using a dataset from the Learning Cell Knowledge Community demonstrate that this method is able to discover learners’ interests effectively. In addition, we find that knowledge interests and collection interests are related and consistent in their subject matter. We further show that learner interest words discovered by the learner-topic model method include learner self-defined interest tags, but reflect a broader range of interests.

Keywords

Open learning environments, Learner interest model, Educational data mining, Learning cell knowledge community, Interaction behaviors, Resource organization

Introduction

Web 2.0 not only brings new ideas and forms of communication for human beings, but also provides opportunities for social interactions focused on knowledge generation, collaborative learning and sharing, and the exchange of knowledge, experience and resources (Kuwara & Richards, 2011). Users’ Web 2.0 interactive content and behavior can offer insight into their learning interests. Detecting user interests based on user-related content and behavior is an important task for enhancing value from online services.

With increasing educational applications of Web 2.0, open learning environments have become an important platform and space for multiple learners to collaboratively create, share and acquire knowledge (Wu & Yu, 2015). Open learning environments involve many different learners, each with different interests that change dynamically. To annotate and manage learner interests, open learning environments usually provide learners with the means to self-define their interests by tagging topics (an example is shown in Figure 1). However, it is difficult for learners to describe their interests in detail. Furthermore, learners will not necessarily update their interest tags as their interests change. In addition, many learners do not actively tag their interests. Thus, it is worth exploring how to discover learner interests automatically in open learning environments.

Learner interests play an important role in Web-based learning environments and are positively related to learning outcomes (Wigfield & Cambria, 2010). Learner interests are reflected in learner generated content and dynamic interactions between the learner and the web-based resources. Essentially, learners express their knowledge interests and knowledge requirements through their online behavior. In open learning environments, the massive amount of data, on both learner generated content and their interactions with online resources, provides opportunities to detect learner interests automatically. At the same time, use of this data enables open learning environments to improve their educational services by adaptively discovering learner needs and automatically recommending relevant resources.

Educational data mining can support the construction of smart learning environments. Mining learner generated content and interaction behaviors to construct a learner interest model is important for offering adaptive learning services in open learning environments. In this paper, we use a text data mining approach to explore the problem of discovering learner interests automatically in open learning environments. Text mining techniques can be employed to mine learner interests from content that they generate and from the text of their online interactions. Topic modelling has attracted recent research attention and been applied in the fields of educational mining and text content analysis. In this study, we present a method to automatically construct a learner interest model in open learning environments using a learner-topic model (LTM). For each learner, interests are composed of two
types: knowledge interests and collection interests. Knowledge interest refers to the learner’s interest in creating content, while collection interest refers to the learner’s interest in collecting content generated by others. We develop an LTM to discover not only learner knowledge interests, but also collection interests.

Figure 1. An example of user self-defined interest tagging

This paper contributes to a better understanding of learner interests in open learning environments. Our study examines different types of learning content data (a) to discover learner knowledge interests and collection interests, (b) to compare the two sets of mined interest data, and (c) to explore the characteristics of and differences between the mined interest data and user self-defined interest tags. Our study aims to answer the following research questions:

**Question 1.** What differences exist between a learner’s knowledge interests and collection interests?

**Question 2.** What differences exist between the mined interest words and self-defined interest tags?

**Question 3.** Is the learner-topic model effective for mining learner interests?

**Literature review**

**Open learning environments**

With the spread of open resources, open learning environments, such as open knowledge communities and massive open online courses, have become increasingly popular (Yang, Qiu, Yu, & Tahir, 2014). Open learning environments provide learners with opportunities for individual knowledge construction, resource annotation, social collaboration, participation and communication (Wu & Yu, 2015). In open learning environments, educational big data are generated from learners’ various interactions and learner generated content. Mining this educational big data in open learning environments is important for offering learners better learning services.

**User interest**

In Web 2.0 environments, user models have attracted increasing attention. User modeling represents several aspects of users such as their knowledge of a subject, their interests, their goals, their backgrounds and other individual traits (Brusilovsky, 2007). Representation of user interests in user models is becoming increasingly popular. Access to user interests makes it easier to satisfy users’ personal needs in recommendation systems,
question answering systems (Ni, Lu, Quan, Liu, & Hua, 2012) and information retrieval and filtering systems (Liu, Chen, Xiong, Ding, & Chen, 2012).

In open learning environments, learners are an important category of users. A learner interest model is a key component of adaptive hypermedia and adaptive educational systems that track learner behaviors and make inferences according to learner interests. Learning behavior actions could offer insight into learners’ topic interest profiles in open learning environments (Peng, Liu, Liu, Gan, & Sun, 2016; Zhao, Cheng, Hong, & Chi, 2015). Knowledge interest and collection interest are reflected in learner generated contents and dynamic learning behavior actions. The smart and personalized educational systems research community has conducted substantial research into the construction of models able to represent user interests (e.g., Zhang, Zhu, Zhao, Gu & Ting, 2008; Gong, 2012; Li, Sagl, Mburu, & Fan, 2016; You, Bhatia, & Luo, 2016). Learner modeling is the process of collecting and computing learning relevant data in educational systems. In the educational environment, analysis of “big data” offers opportunities for constructing such learner interest models.

Text mining in eLearning

Mining of educational data can play a supportive role in eLearning. Hwang (2005) proposed a data mining approach to assist teachers in providing information tailored to guide individual students in their learning efforts. In recent years, text mining has become popular in educational data mining applications. Text mining aims to find and extract useful, latent or interesting patterns and models from unstructured text documents. Text mining can be used to identify, extract, integrate, and exploit knowledge for eLearning efficiently and effectively (He, 2013). In recent years, a number of studies have used text mining techniques to analyze learning-related data. For example, text mining techniques were used to automatically analyze data from online questions, interactions and chat messages and predict final student grades (Abdous, He, & Yen, 2012; He, 2013). Other studies analyzed student attitudes towards learning through mining lecture data, and explored correlations between learning attitudes and learning achievement through analyzing the texts of student answers to a questionnaire (Minami & Ohura, 2013; Minami & Ohura, 2015). Goda and Mine (2011) estimated learning situations based on text mining of student comments.

Topic modeling is a text mining method for estimating topics in documents and clustering documents based on latent topics. Sekiya, Matsuda and Yamaguchi (2010) used latent topic modeling to analyze course syllabi. Sorour, Goda and Mine (2015) used two types of machine learning techniques to learn the relationships between comment data analyzed by Latent Semantic Analysis (LSA) and final student grades. Sorour, Goda and Mine (2017) applied Latent Dirichlet Allocation (LDA) and Probabilistic Latent Semantic Analysis (pLSA) to predict student grades in each lesson through mining student comment data.

There have also been several studies using topic modeling focused on modeling users and mining user interests in Web 2.0 environments such as Microblogs and Twitter. Pennacchiotti and Popescu (2011) applied topic modeling techniques to classify users by considering user profiles, behaviors, content and social network features. Ishii, Mizoguchi, Kimita and Shimomura (2015) proposed a topic model for clustering learners based on educational counseling content. Michelson and Maucksassy (2010) discovered topics of interest for Twitter users based on their posts. Xu, Ru, Xiang and Yang (2011) proposed a method for discovering an author’s interest on Twitter with a twitter-user model. Collectively, these efforts demonstrate that content features are highly valuable, in general, and that topic modeling techniques are reliable and effective for social media user classification.

Currently, in the eLearning field, there are fewer studies using topic modeling to mine learner interests. For example, Zhang, Zhu, Zhao, Gu and Ting (2008) used behavioral analysis for interest mining in e-learning. Tobarra, Robles-Gómez, Ros, Hernández and Caminero (2014) analyzed student behaviors and relevant topics in virtual learning communities. Peng, Liu, Liu, Gan and Sun (2016) explored learners’ topic interests by mining course reviews using an LDA-like model, showing that learner interactions with these texts were helpful in building learner topic interest profiles; moreover, the combination of interactive behavioral features with textual content was useful for mining learner topic interests (Peng, Liu, Liu, Gan, & Sun, 2016). To sum up, learners not only create content that interests them. They are also more likely to collect content created by other learners, which interests them. Consequently, in this study, we use the topic model mining approach to discover learner interests automatically by integrating learner generated content and data about their interactions with resources available in open learning environments. We propose a method to mine learner interests automatically in open learning environments using a learner-topic model.
Methodology

We introduce a LTM for mining learner interests based on two types of learning-related data. An overview of our methodology is shown in Figure 2. In this section, we first describe the collection of the two types of learning-related data, which includes the learner creation data and the collection data. Next, we briefly review the LDA model. Then based on LDA, we introduce our learner-topic model and demonstrate how to mine learner interests with our LTM.

![Figure 2. A framework of the methodology](image)

Learning data collection

Learning Cell is a resource organization model for ubiquitous learning in a seamless learning space (Yu, Yang, Cheng, & Wang, 2015). The Learning Cell Knowledge Community (LCKC) (see http://lcell.bnu.edu.cn), inaugurated in May 2011, is an open knowledge community constructed based on Learning Cell (Yang, Qiu, Yu, & Tahir, 2014). As of April 20, 2017, LCKC had 24508 registered users and 81218 learning cells.

Our study uses learning-related data from the learning resource database of LCKC. For privacy protection, we removed learners’ real names from the dataset. Before using topic modeling for interest mining, we first extract learning content data and obtain the sets of created learning content data and collected learning content data for each learner. An example of learner creation data and collection data for a single LCKC learner is shown in Figure 3.
Data preparation

We use the natural language segmentation system of IKAnalyzer (see https://www.oschina.net/p/ikanalyzer) to segment the Chinese text obtained from LCKC and extract Chinese words. In our study, we use the stop word dictionary to filter out stop words.

LDA model

The primary goal of this study is to develop a model to mine the learning-related data for learner interests. The LDA topic model (Blei, Ng, & Jordan, 2003) is a general Bayesian probabilistic framework for modeling documents linked by a layer of latent topics. It is a method for clustering the documents based on latent topics. The LDA topic model assumes that words in each document were generated from a mixture of latent topics, where each latent topic is represented as a multinomial probability distribution over words. A document is modeled as a set of draws from a mixture distribution over a set of latent topics and a topic is modeled as a probability distribution over words. There are many methodological extensions to LDA. LDA topic models have been applied and have demonstrated reliability in many text mining tasks. The following learner-topic model builds on LDA and its extensions.

Learner-Topic model

In open learning environments, learners not only create learning content, but also collect learning content created by other learners in which they are interested. These activities suggest that the generative process for learning content should meet the following rules for a topic model:

- Learning content created and collected by a learner are related to the learner’s interests. The learning content originates from the learner’s topic distribution. In the generative process of learning content, we should choose a latent topic from the learner’s topic distribution for each word in the learning content.
- Learner interest is composed of two parts: knowledge interest and collection interest. Knowledge interest refers to the learner’s interest in creating learning content. Collection interest refers to the learner’s interest in collecting learning content.
- Learning content created by a learner should be generated from the learner’s knowledge topic distribution, while learning content collected by a learner should be generated from the collection topic distribution.

Based on the above rules, we propose an LTM for topic mining in open learning environments. The LTM models the generation process of the learner’s learning content of interest, and then we deduce learner interest from the model. For each learner, learning content of interest is composed of (1) created learning content and (2) collected learning content. A learner creates learning content based on his/her knowledge interests, and therefore the topic distribution of the created learning content should be determined by the learner’s knowledge interests. Thus, we...
can model the generation process for created learning content based on the learner’s knowledge interests. Collected learning content is created by other learners. The current learner collects that learning content, which reflects the learner’s collection interests. Therefore, we can derive the learner’s collection interests by aggregating topic distributions of the collected learning content. In our proposed learner-topic model, we aggregate all the learning content created by a learner as a single document $D_{\text{created}}$ and all the learning content collected by a learner as a single document $D_{\text{collected}}$. Thus, each document essentially corresponds to a learner.

In open learning environments, each learner can create learning content and collect other learners’ learning content. Therefore, we can model the generation process of all the learning content for each learner by considering learner interests in terms of knowledge interests and collection interests. This is the reason why we divide learner interest into two parts in our model. We can get the topics of knowledge interest and the topics of collection interest for a particular learner by learner-topic modeling. Because learner interests consist of two parts, we can generate the learner interest words by aggregating knowledge interests and collection interests.

**Learner knowledge interest model**

For a learner, the generating probability of the word $w$ is given as follows:

$$P(w|l, \theta, \phi) = \sum_{z=1}^{K} P(w|z, \phi_{z})P(z|l, \theta_{l})$$  \hspace{1cm} (1)

where the document is created by learner $l$, $z$ is a topic, $K$ is all topics, $\phi_{z}$ is the multinomial distribution of the learner $l$ over topics, $\phi_{z}$ is the multinomial distribution of the topic $z$ over words.

The generation process algorithm for the Learner Knowledge Interest Model is as follows:

1. for each topic $z \in [1, K]$ do
2. choose a distribution over word $\phi_{z}$ from a Dirichlet distribution with parameter $\beta$.
3. end for

4. for each learner $l \in [1, L]$ do
5. choose a distribution over topics $\theta_{l}$ from a Dirichlet distribution with parameter $\alpha$.
6. for the $n$th token in the document set $D_{\text{created}}$ created by $l$ do
7. choose a latent topic $z_{ln}$ from the multinomial distribution $\theta_{l}$.
8. generate the word token $w_{ln}$ from the multinomial distribution $\phi_{z_{ln}}$.
9. end for
10. end for

**Learner collection interest model**

For a learner, the generating probability of the word $w'$ is given as follows:

$$P(w'|l, \theta', \phi') = \sum_{z'=1}^{K} P(w'|z', \phi'_{z'})P(z'|l, \theta'_{l})$$  \hspace{1cm} (2)

where the document is collected by learner $l$, $z'$ is a topic, $K$ is all topics, $\phi'_{z'}$ is the multinomial distribution of the learner $l$ over topics, $\phi'_{z'}$ is the multinomial distribution of the topic $z'$ over words.

The generation process algorithm of the Learner Collection Interest Model is as follows:

1. for each topic $z' \in [1, K]$ do
2. choose a distribution over word $\phi'_{z'}$ from a Dirichlet distribution with parameter $\beta'$.
\[ \mathcal{O}'_{z'_{in}} \sim Dir(\beta') \]

3. **end for**
4. **for** each learner \( l \in [1, L] \) **do**
5. choose a distribution over topics \( \mathcal{O}'_{i} \) from a Dirichlet distribution with parameter \( \alpha' \).
6. **end for**
7. for the \( n \)th token in the document set \( D_{\text{collected}} \) collected by \( l \) **do**
8. choose a latent topic \( z'_{in} \) from the multinomial distribution \( \mathcal{O}'_{i} \).
9. generate the word token \( w'_{in} \) from the multinomial distribution \( \mathcal{O}'_{z'_{in}} \).
10. **end for**
11. **end for**

---

**Discovering learner interests**

We can find the word proportions over each topic and extract the representative words for each latent topic. Then we can find the latent topic proportions over the created learning content and extract the knowledge interests for each learner. We can also get the latent topic proportions over the collection learning content and extract the collection interests for each learner. The learner-topic model thus mines learner interests from two aspects: knowledge interests and collection interests. We use the java open source software of Machine Learning for Language Toolkit named Mallet (see http://mallet.cs.umass.edu) to implement the learner-topic model.

**Results**

**Dataset**

For our evaluation, we used a text dataset extracted from the Learning Cell Knowledge Community consisting of learning cell documents and learners. Stop words were removed from each learning cell document. In the learner-topic model generation phase, we selected 3538 learners from the Learning Cell Knowledge Community, each of whom had created at least one Learning Cell or collected one Learning Cell. This selection process yielded 45512 Learning Cells linked to these learners.

**Parameters estimation**

The learner-topic model requires specification of the Dirichlet prior hyper parameters \( \alpha, \beta, \alpha', \beta' \). According to a previous study (Blei, Ng & Jordan, 2003), the Dirichlet prior hyper parameters setup of the learner-topic model in the experiment is: \( \alpha = 50/K; \beta = 0.01; \alpha' = 50/K; \beta' = 0.01 \). In our research, we used the perplexity (Blei, Ng & Jordan, 2003) to choose the optimal \( K \) (number of topics). Latent topics were extracted from a single sample using the 2000th iteration of Gibbs sampling.

**Learner knowledge interest topics**

In Table 1, we list knowledge interest topics extracted by the learner-topic model for five learners, including the top ten words of each topic. These words appear with high probability in each topic. The specific meaning of each topic is based on analysis of the semantics of the representative words. From Table 1, it is easy to confirm that the top ten high-frequency words for each topic are closely related to that topic.

<table>
<thead>
<tr>
<th>Learner</th>
<th>Topic</th>
<th>Top ten most frequent words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learner1</td>
<td>“Educational technology”</td>
<td>education (教育) informatization(信息化) integration(整合) information technology(信息技术) learning(学习) curriculum(课程) teacher(教师)</td>
</tr>
<tr>
<td></td>
<td>“Digital education”</td>
<td>photography(摄影) report(报告) research(研究) education(教育) project(课题) frontier(前沿) digitization(数字化) culture(文化) module(模块) West(西区)</td>
</tr>
</tbody>
</table>

198
### Learner collection interest topics

In Table 2, we list collection interest topics extracted by the learner-topic model for five learners, including the top ten words of each topic. Again, we analyze the semantics of the representative words to get the specific meaning of each topic. We find that the top ten high-frequency words for each topic are closely related to the topic.

<table>
<thead>
<tr>
<th>Learner</th>
<th>Topic</th>
<th>Top ten most frequent words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learner1</td>
<td>“Educational technology”</td>
<td>educational technology(教育技术) network(网络) paradigm(范式) new solutions(新解) change(变迁) basic problem(基本问题) instructional design(教学设计) teaching(教学) society(社会) learning theory(学习理论)</td>
</tr>
<tr>
<td></td>
<td>“Educational technology in society”</td>
<td>learning cell(学习元) platform(平台) summary(概述) enterprise(企业) university(大学) method(方法) advice(意见) improvement(改进) usage(使用) learning(学习) group(小组)</td>
</tr>
<tr>
<td>Learner2</td>
<td>“Poetry and computer program”</td>
<td>employment(就业) common problem(常见问题) poetry(诗词) ancient(古代) interpret(解读) program(编程) notice(公告) editing(编辑) server(服务器) leapfrogging(跨越式)</td>
</tr>
<tr>
<td></td>
<td>“Educational technology”</td>
<td>educational technology(教育技术) network(网络) paradigm(范式) new solutions(新解) change(变迁) basic problem(基本问题) instructional design(教学设计) teaching(教学) society(社会) learning theory(学习理论)</td>
</tr>
<tr>
<td>Learner3</td>
<td>“Computer assisted teaching”</td>
<td>Chongwen(崇文) Minsheng(民生) experimental school(实验学校) strategy(策略) extension(推广) happy holiday(快乐的节日) Heilongjiang(黑龙江省) computer(电脑) child(幼儿) experience(心得体会)</td>
</tr>
<tr>
<td></td>
<td>“Educational technology”</td>
<td>educational technology(教育技术) network(网络) paradigm(范式) new solutions(新解) change(变迁) basic problem(基本问题) instructional design(教学设计) teaching(教学) society(社会) learning theory(学习理论)</td>
</tr>
</tbody>
</table>
Learner interest words and self-defined tags

For this part of the study, we chose learners who had self-defined interest tags, enabling us to evaluate the model by comparing learner interests with learner self-defined interest tags. Learner self-defined interest tags are sets of keywords defined by a learner and used to describe his/her specialties and interests in LCKC (as shown in Figure 1). Therefore, learner self-defined interest tags are an informal reflection of learner interests.

To study the difference between learner interest words and learner self-defined interest tags, we first combined knowledge interest words (ten words selected from two of the most popular knowledge interest topics) and collection interest words (ten words selected from two of the most popular collection interest topics) as interest words for each learner. We then compared the learner self-defined interest tags with the combined set of learner interest words as shown in Table 3. For example, Learner3 provides the tag words “Chinese course(语文),” “English course(英语),” “leapfrogging (跨越式)” and “teaching(教学).” Therefore, we might expect that Learner3 is also interested in the research of the “leapfrogging project (跨越式项目).”

<table>
<thead>
<tr>
<th>Learner</th>
<th>Interest tags</th>
<th>Interest words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learner1</td>
<td>educational technology(教育技术学) ubiquitous learning(泛在学习) web based education platform(网络教育平台) learning cell(学习元) leapfrogging project(跨越式项目) poetry(诗歌)</td>
<td>education(教育) informatization(信息化) integration(整合) information technology(信息技术) learning(学习) photography(摄影) report(报告) research(研究) educational technology(教育技术) network(网络) Paradigm(范式) new solutions(新解) change(变迁) learning cell(学习元) platform(平台) summary(概述) enterprise(企业) university(大学) method(方法)</td>
</tr>
<tr>
<td>Learner2</td>
<td>educational technology(教育技术学) ubiquitous learning(泛在学习) web based education platform(网络教育平台) leapfrogging project(跨越式项目)</td>
<td>learning(学习) design(设计) mobile(移动) resource(资源) development(开发) method(方法) problem(问题) file(文件) introduction(入门) database(数据库) command(命令) employment(就业) common problem(常见问题) poetry(诗词) ancient(古代) interpret(解读) program(编程) educational technology(教育技术) network(网络) paradigm(范式) new solutions(新解) change(变迁)</td>
</tr>
<tr>
<td>Learner3</td>
<td>leapfrogging project(跨越式项目)traveling(旅行) English course(英语) Chinese course(语文)</td>
<td>Chinese course(语文) instance(案例) learning cell(学习元) train(培训) share(分享) teaching(教学) photography(摄影) report(报告) research(研究) educational technology(教育技术) project(课题) Chongwen(崇文) Minsheng(民生) experimental school(实验学校) leapfrogging(跨越式) English course(英语) educational technology(教育技术) network(网络) paradigm(范式) new solutions(新解) change(变迁)</td>
</tr>
</tbody>
</table>
User evaluation

We invited 25 LCKC users to participate in the experiment. Each user evaluated their generated knowledge interests, collection interests and the overall learner interest result quantitatively, by counting the number of the discovered top ten words that accurately reflected their interests. The precision is equal to the number of the discovered top ten words relevant to an individual’s interests divided by ten. Similar metric has been used in evaluating tasks (Michelson & Macskassy, 2010). Cronbach’s alpha (.701) indicates an acceptable internal consistency estimate of reliability of evaluation scores.

<table>
<thead>
<tr>
<th>Interest</th>
<th>Mean (average precision)</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge Interest</td>
<td>.820</td>
<td>.076</td>
<td>25</td>
</tr>
<tr>
<td>Collection Interest</td>
<td>.764</td>
<td>.075</td>
<td>25</td>
</tr>
<tr>
<td>Learner Interest</td>
<td>.884</td>
<td>.068</td>
<td>25</td>
</tr>
</tbody>
</table>

The average levels of precision for knowledge interests, collection interests and overall learner interests are presented in Table 4. The percentages of precision for knowledge interests, collection interests and overall learner interests are presented in Figure 4.

![Precision Percentage](image.png)

Discussion

This study uses a learning content mining approach to automatically discover learner interests, using the combination of content generated by each learner and the dynamic interactions of learners with the online
resources to reflect learner interests. In this section, we discuss the results of our three research questions. In considering the learner knowledge and collection interests extracted in the experiment, we judge and analyze their relationships and differences from each learner’s self-defined interest tags through observation.

**Differences between learner knowledge interests and collection interests**

In comparing the result of extracted knowledge interests (Table 1) with collection interests (Table 2), we note that, for each learner, knowledge interests and collection interests are similar. For example, “educational technology” is present in both the created learning content and the collected learning content for Learner1. Similarly, “computer techniques and educational technology” feature in both the created learning content and the collected learning content for learner 2. Visual analysis of the differences between knowledge interests and collection interests for each user suggests that these areas of interest are related and consistent in their subject matter. The results shown in Tables 1 and 2 address the research question regarding similarities and differences between learner knowledge interests and collection interests.

**Differences between the mined interest words and self-defined interest tags**

In comparing the result of mined interest words with self-defined interest tags (Table 3), we find that, for each learner, the set of mined interest words contains the learner’s self-defined interest tags. For example, Learner1’s self-defined tag of “educational technology” is among the interest topics (topic of “educational technology” and topic of “educational technology in society”) extracted for Learner1. Similarly, Learner2’s self-defined tag of “ubiquitous learning” is contained in the interest topics (topic of “mobile learning resources” and topic of “educational technology”) elicited for Learner2. The results shown in Table 3 address the research question regarding the relationship between mined interest words and learner self-defined interest tags. We find that, for a particular learner, the set of mined interest words contains the learner’s self-defined interest tags, but is not limited to the scope of the learner’s self-defined interest tags. For example, Learner4 is also interested in the topic of “architecture,” which is not found in Learner4’s self-defined interest tags.

**Effectiveness of learner-topic model**

Experimental results show that the mean value of the learner interests is higher than the mean value of knowledge interests and collection interests separately. Generating learner interests from the knowledge interests and collection interests to acquire learners’ interest characteristics is more accurately. User experimental results in Table 4 and Figure 4 indicate that the learner-interest model for mining learner interests is appropriate and effective, and support the use of the topic-modeling approach for discovering learner interests in open learning environments. In comparing user self-defined interest tagging with the results of the learner-topic model, it suggests that the learner-topic model represented in Figure 5 not only automatically discovers learner interests in detail, but also can be used to update learner interest tags when their interests change. This model thus enables learners to annotate their interests dynamically without manual intervention.

![Figure 5. Feature of the learner-topic model in LCKC system](image-url)
Conclusions

Building a learner interest model has been an important research topic in open learning environments. This paper discussed and demonstrated a method of using a LTM to solve the problem of mining learner interests and automatically generating learner interests in open learning environments. Experiments on a dataset from the Learning Cell Knowledge Community demonstrate that the method is able to discover learner interests effectively. Moreover, we find that knowledge interests and collection interests are related and consistent in their subject matter. Finally, the learner interest words discovered by the LTM method reflect self-defined interest tags, but cover a broader range of interests. In open learning environments, different interactions between learners and various sources of content indicate learners’ attention to diverse topics of information. Learners’ varied interests are revealed by the combination of knowledge interests and collection interests, which we derived from learner generated content plus data on the dynamic interactions between learners and online resources. Different semantic behavior actions such as “create,” “like” could also provide insight into online learners’ topic interest profiles (Peng, Liu, Liu, Gan, & Sun, 2016; Zhao, Cheng, Hong, & Chi, 2015). Our findings offer an approach to automatically discover learner interests in detail from learner generated content and dynamic interactions and relationships in open learning environments. Learning resource content and learners’ behavioral features (“create” and “collect”) are merged to more accurately acquire the learners’ interests. Other useful operational behaviors will be utilized in the process of educational data mining into the learner-topic model, e.g., “comment,” “share.”

Further research could be conducted to address additional relationships between learners and online resources of interest. In our future work, we plan to use the mined interest words for personal resource recommendation, learning peers discovery and experts finding.

Acknowledgements

This research is funded by the project “the Research on Internet plus Educational System” (Project No. 16JZD043), major research subject of philosophy and social science of the ministry of education.

References


Minami, T., & Ohura, Y. (2013). Lecture data analysis towards to know how the students’ attitudes affect to their evaluations. In *Proceedings of the 8th International Conference on Information Technology and Applications* (pp. 164-169). Sydney, Australia: ICTIA.


A Learning Analytics Approach to Investigating Factors Affecting EFL Students’ Oral Performance in a Flipped Classroom

Chi-Jen Lin and Gwo-Jen Hwang

Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology.
chi.jen1023@gmail.com // gjhwang.academic@gmail.com

ABSTRACT

Flipped classrooms have been widely adopted and discussed by school teachers and researchers in the past decade. However, few studies have been conducted to formally evaluate the effectiveness of flipped classrooms in terms of improving EFL students’ English oral presentation, not to mention investigating factors affecting their flipped learning outcomes. In this study, an online community-based flipped learning approach was proposed for an EFL oral presentation course; moreover, a learning analytics approach was used to analyze factors affecting the students’ oral presentation outcomes. An 18-week research design was implemented with the online community-based flipped classroom using Facebook as the platform for facilitating and recording peer-to-peer interactions during the flipped learning process. In addition, the students’ learning performance and perceptions were collected in 3 learning stages during the 18 weeks. The experimental results reveal positive effects of the online community-based flipped instruction over the conventional video-based instruction. That is, first, the online community-based flipped instruction using mobile devices can enhance students’ English oral performance. Moreover, it was found that the high improvers had a significantly higher frequency of online participation, as well as more interactive behaviors and greater satisfaction with the flipped classroom than the low improvers. These results imply that the online community flipped classroom could not only provide learning materials and out-of-class learning for students, but could also help them become more responsible and autonomous in their learning and communication. These findings could be valuable references for those who intend to conduct effective flipped classrooms with an online community to facilitate students’ before-class learning participation and to improve their in-class learning performance.

Keywords

Flipped classroom, Mobile language learning, English oral performance, Learning analytics

Introduction

The significant rise in flipped teaching and learning in education (Bergmann & Sams, 2012) has not only changed the traditional learning patterns and teacher-centered learning modes, but has also improved students’ learning achievement and increased the interaction among learners and teachers (Hwang & Lai, 2017). Flipped learning has been recognized as a pedagogical approach to integrating instructional videos for students to self-learn outside of class by watching and reviewing the learning content or pre-class assignments before class (DeLozier & Rhodes, 2017). This learning mode may engage students in meaningful learner-to-learner or learner-to-teacher interactions in the community (Schultz, Duffield, Rasmussen, & Wageman, 2014). In flipped learning environments, there is more time to help students better prepare for and engage in learning activities or participate in classroom lectures (Bergmann & Sams, 2015; Cockrum, 2013), such as group project-based learning, in-depth discussion, or mobile technology-enhanced learning (Hwang, Lai, & Wang, 2015; Hwang & Lai, 2017).

Consequently, investigating factors affecting students’ language learning in flipped classrooms is a critical task for higher education institutions. In the past few years, various studies have been conducted to engage students in the flipped classroom approach in English language learning (Ahmed, 2016; Al-Harbi & Alshumaimeri, 2016; Asoodar, Atai, Vaezi, & Marandi, 2014; Basal, 2015; Han, 2015; Sam, 2016; Soliman, 2016; Stockwell, 2013; Zhang, Du, Yuan, & Zhang, 2016). Flipped teaching and learning involve regular and systematic use of interactive technologies in the learning process (Al-Zahrani, 2015).

This is an area that is of interest to researchers, teachers, material writers, and application developers in the digital world. Researchers have indicated that the improvement in students’ English learning performance, including listening, speaking, reading, and writing skills, is related to their preparation, participation, attitudes, learner autonomy, sense of community, collaboration, or different learning experiences (Ahmed, 2016; Al-Harbi & Alshumaimeri, 2016; Asoodar et al., 2014; Basal, 2015; Han, 2015; Sam, 2016; Soliman, 2016; Stockwell, 2013; Zhang et al., 2016). In addition, some researchers have attempted to improve students’ English speaking skills through video blogging, blogs or multimodal video technology as a learning instrument for improving English-speaking performance (Hung, 2016; Hung & Huang, 2015).
On the other hand, the individual learning process and learning traces play a central role in instructional management and in teaching development (Graf, Yang, Liu, & Kinshuk, 2009). Each learner has individual learning behaviors due to different learning abilities, attitudes, motivation, and so on. However, only a few scholars have investigated factors in flipped classrooms or have used a learning analytics approach to analyzing students’ feedback patterns in English language teaching.

In this study, an online community-based flipped learning approach was implemented using mobile technologies to facilitate students to engage in out-of-class learning as well as in-class activities of an English course. A particular strength of this study is that it focuses on the students’ English oral performance, participation, and feedback patterns using mobile technologies in the online community-based flipped classroom. An experiment was conducted using Oral-Aural Drills in an English course in my university to evaluate the consequences and effectiveness of the proposed approach, with the aim of answering the following research questions:

- Can the online community-based flipped learning approach improve the students’ oral performance in comparison with the conventional video-based learning?
- Does the online community-based flipped learning approach impact students’ participation, and is there a difference between high improvers and low improvers?
- Does the online community-based flipped learning approach impact students’ feedback patterns, and is there a difference between high improvers and low improvers?
- What are high improvers’ and low improvers’ perceptions of implementing the online community flipped classroom?

**Literature review**

The flipped classroom has recently been recognized as an effective learning approach in various courses (Lai & Hwang, 2016; Hwang et al., 2015). The adoption of flipped classrooms in ELT (Sam, 2016; Soliman, 2016) not only helps teachers reach students with different abilities or learning achievements, but also improves classroom management, giving teachers more time to interact with each student (Basal, 2015; Bergmann & Sams, 2015; Sung, 2015). In other words, flipping the classroom benefits students in various aspects, including enhancing their creative thinking (Al-Zahrani, 2015), listening comprehension (Ahmad, 2016), grammar skills (Al-Harbi & Alshumaimeri, 2016), reading comprehension (Huang & Hong, 2016), writing skills (Ahmed, 2016), English pronunciation (Zhang et al., 2016), and overall English proficiency (Wu, Hsieh, & Yang, 2017). The flipped classroom also helps students become more responsible for their learning (Han, 2015; Sung, 2015), and allows teachers more individual interaction with every student and the ability to develop better relationships with all their students (Zhang & Wu, 2016). Moreover, flipping the instruction significantly reduces negative behavior in the classroom (Cockrum, 2013).

In the past decade, many scholars and educators have engaged students in English language flipped classrooms and have distilled ideas in the flipped learning world to help more teachers chart a path towards pedagogical innovation. Conversation, public speaking, English speech, and presentation are essential components of English as a foreign language speaking skills (Koçak, 2010; Swain, 1985). Cockrum (2013) indicated that students can produce successful comprehensive speech or presentations in three ways. First, they can write effectively to convey the message. Second, teachers have students deliver a speech with appropriate enunciation, body language, gestures, volume, and demeanor. Third, students make several formal and informal presentations to practice these skills. In addition, Bergmann and Sams (2015) stated that various types of hands-on activities, such as working with different media, script writing, rehearsals, rewrites, and movie filming in the learning process allow students to engage more deeply and motivate them to learn, while also encouraging their creativity and understanding.

Consequently, teachers could use a flipped speaking course to increase students’ motivation and engagement to help them take more responsibility for their learning (Bergmann & Sams, 2015). Particularly in recent years, several studies have been conducted on flipping English language teaching. One of the examples is Hsieh, Huang, and Wu’s (2017) series of studies regarding flipped classrooms. They indicated the effectiveness of flipped classrooms in terms of enhancing students’ oral proficiency, and showed that the application of technologies can facilitate English language teaching. Their results also revealed that the students were satisfied with the flipped classroom, accepted the technology, and were motivated by the incorporation of mobile learning. They further reported that the students learning with the mobile-based flipped learning approach had better learning outcomes and attitude for active and continuous learning than those learning with the conventional lecture-based approach. However, in their study, students’ learning behaviors were not investigated to support the findings. Therefore, in the present study, an online community-based flipped classroom was...
proposed and a learning analytic approach (Agudo-Peregrina, Iglesias-Pradas, Conde-González, & Hernández-García, 2014; Slade, & Prinsloo, 2013) was used to analyze factors and students’ interactive behaviors affecting students’ oral proficiency.

Online community-based flipped learning

In this study, the popular social networking site, Facebook, was adopted as the platform since all of the participants had Facebook accounts, and most of them used it frequently. It only required minimal computer skills and provided three essential functions: a wall, newsfeed, and live stream video. The teacher selected educational videos related to the target language from TED, VoiceTube, and YouTube and uploaded them to the wall. The students could view one or two videos each week in the newsfeed and respond to the designed open questions related to the video in comments area, as shown in Figure 1.

Experimental group (N=33) Out-of-class learning on Facebook

Outside the classroom, students could use their computer, iPad, mobile phone, or smartphone to watch the video and respond to the questions. If they did not complete the out-of-class learning (i.e., watch the video(s) and do the oral or written assignments on the newsfeed), they would be asked to complete it before the lecture started. The final function of the live stream video on Facebook was for the students to record dialogues and share their videos. In addition, they could also give comments and get others’ feedback immediately, as shown in Figure 2. Therefore, each of the students in the experimental group was required to join the Facebook group for this class.

Figure 1. Facebook as a platform for students’ online community-based flipped learning
Research design

Context of the study

This research was conducted in 2017 as part of two Oral-Aural Drills in English courses in a Taiwanese university. The classes met for 2 hours per week, 18 weeks per semester, were taught by the same English lecturer, and shared the same objectives: (1) to enhance students’ English oral performance on the given topics and (2) to promote students’ English expression and oral presentation skills.

Participants

The study adopted a quasi-experimental design in which two classes of ELF (English as a Foreign Language) students were assigned to an experimental group and a control group. The experimental group with 33 students adopted the online community-based flipped classroom approach, while the control group with 16 students learned with the conventional video-based learning approach. The students in these two classes were freshmen who took the placement test set by the Language Center at the University and had the same level of English proficiency; that is, their test scores ranged from 350-550, which, according to the TOEIC official information, places them at the level of elementary proficiency plus, that is, they had the basic competence required for starting face to face conversations in English. The average age of the students was 18.

Instruments

We collected four types of data in this study: three oral performance video clips, students’ participation, students’ interactive behavior, and the survey questionnaire of the students’ perceptions of the flipped classroom. The level of the English oral tests was determined by the English lecturers in the Language Center at the University, and two English experts selected the oral topics from the elementary proficiency plus level and then assigned the topics to the Time 1 oral test, the Time 2 oral test, and the Time 3 oral test, indicating that the level
of difficulty of the three oral tests was the same. Within the 18 weeks, the students’ three video clips of English oral performance were uploaded to Facebook. The following section describes the rubric of English oral performance.

The rubric of English oral performance

The rubric for measuring the students’ English Oral Performance was developed by DePalma, Cartland, and Neumaier (2013). Table 1 shows the rubric, which consists of six dimensions with a total score of 24 points, 4 points for each dimension, that is accuracy, comprehensibility and pronunciation, fluency, comprehension, content, and maturity of language. These six dimensions measure how well the students used sentence structures, vocabulary, and correct grammar; how well they communicated ideas and used correct pronunciation with no significant errors; how well they understood all verbal cues and responded appropriately; whether their speaking content included information such as an opening, body, and conclusion; and whether they could communicate effectively by using appropriate words, expressions, eye contact, and gestures. The aim of this study was to develop an understanding of the experimental group and control group students’ English-speaking learning activities and practices. As a result, the students would be able to use English to communicate, present themselves, and engage in a real-world community.

Table 1. The rubric of English oral performance

<table>
<thead>
<tr>
<th>Scores</th>
<th>Accuracy</th>
<th>Comprehensibility and Pronunciation</th>
<th>Fluency</th>
<th>Comprehension</th>
<th>Content</th>
<th>Maturity of the Language</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ability to use sentence structure, vocabulary, and grammar correctly with no significant errors</td>
<td>Ability to communicate ideas and be understood using correct pronunciation with no significant errors</td>
<td>Ability to communicate clearly and smoothly with only natural hesitation</td>
<td>Ability to understand all verbal cues and always respond appropriately</td>
<td>Inclusion of all required information, including opening, body, and conclusion</td>
<td>Inclusion of details beyond the minimum requirements (ex: using words/expression/eye contact/gesture)</td>
</tr>
<tr>
<td>4</td>
<td>Ability to use sentence structure, vocabulary, and grammar correctly with minimal errors</td>
<td>Ability to communicate ideas and be understood using correct pronunciation with no minimal errors</td>
<td>Ability to communicate clearly and smoothly with minimal hesitation</td>
<td>Ability to understand most verbal cues and almost respond appropriately</td>
<td>Inclusion of most required information</td>
<td>Inclusion of details beyond the minimum requirements</td>
</tr>
<tr>
<td>3</td>
<td>Ability to use sentence structure, vocabulary, and grammar correctly with some errors</td>
<td>Ability to communicate ideas and be understood using correct pronunciation with some errors</td>
<td>Ability to communicate with some prompts</td>
<td>Ability to understand some verbal cues and sometimes requires prompts</td>
<td>Inclusion of some required information</td>
<td>Inclusion of details beyond the minimum requirements</td>
</tr>
<tr>
<td>2</td>
<td>Inability to use sentence structure, vocabulary, and grammar correctly (many errors)</td>
<td>Inability to communicate ideas and be understood (many errors in pronunciation)</td>
<td>Inability to communicate ideas unless given prompts</td>
<td>Inability to understand verbal cues and respond appropriately</td>
<td>Inclusion of little no required information</td>
<td>Good Luck!</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Coding scheme for assessing students’ interactive behaviors

To explore the students’ oral interactive patterns in the online community-based flipped classroom, a coding scheme was developed to code their behaviors. A model of interactive behaviors in an EAP (English for
Academic Purposes) classroom, originally developed by Unlu and Wharton (2015), was administered. The concept labels were used to generate a list of concepts representing all the feedback from students in their responses on Facebook. Two experts were invited to confirm the suitability of the codes and the corresponding definitions, and the accuracy of the coding result based on the coding scheme, as shown in Table 2. After the learning activity, two researchers were asked to code the students’ online interactive behaviors based on the coding scheme. For those inconsistent coding results, a judge discussed with the two researchers and came out with the final result of full agreement among all.

Table 2. The coding scheme of students’ interactive behaviors on Facebook

<table>
<thead>
<tr>
<th>Category</th>
<th>Code</th>
<th>Definition</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student feedback</td>
<td>CL</td>
<td>Clarification</td>
<td>Students’ attempts to explain reasons</td>
<td>I think Vincent Chao can be so successful because he has studied abroad and he is very brave to deal with many challenges.</td>
</tr>
<tr>
<td></td>
<td>CO</td>
<td>Confirmation</td>
<td>Students’ validation to the significance of ideas</td>
<td>Well, I agree that the speaking style comes a lot from listening to speeches over and over again such as from American politicians.</td>
</tr>
<tr>
<td></td>
<td>RE</td>
<td>Retelling</td>
<td>Students retell the sentences</td>
<td>“It is ok to step out and use a language even though you may not feel 100% confident in it,” said Vincent Chao.</td>
</tr>
<tr>
<td></td>
<td>CH</td>
<td>Challenge</td>
<td>Students’ response to the idea with some level of disagreement</td>
<td>If you had not studied hard in your childhood, you would not be a reporter. A man who cannot speak Chinese can be a reporter in Taiwan? BTW, Vincent speaks too fast.</td>
</tr>
<tr>
<td></td>
<td>SU</td>
<td>Suggestion</td>
<td>Students offering possible ideas or suggestions</td>
<td>The best way to study a language is to use it: to speak, to write, to listen and to read.</td>
</tr>
<tr>
<td></td>
<td>SR</td>
<td>Surmise</td>
<td>Students guess something</td>
<td>Does he speak better than me? Yeah</td>
</tr>
</tbody>
</table>

Questionnaire of students’ perceptions of the online community-based flipped classroom

The questionnaire of perceptions of flipped learning was modified based on the survey developed by Al-Zahrani (2015). It used a 5-point rating scheme (5 = “strongly agree” and 1 = “strongly disagree”) to evaluate the students’ views on the flipped classroom with the following 14 items:

1. The flipped classroom offers me the opportunity to review the lectures as many times as I need to.
2. The flipped classroom offers me access to the online course tools and materials.
3. The flipped classroom helps me to use various e-learning resources.
4. The flipped classroom helps me to enrich my learning experience.
5. The flipped classroom helps me to connect theory with practice in real life.
6. The flipped classroom helps me to effectively cooperate with my classmates.
7. The flipped classroom facilitates more communication between me and my teacher.
8. The flipped classroom facilitates more communication between me and my classmates.
9. The flipped classroom helps me to effectively participate in the learning activities.
10. The flipped classroom enables me to manage my learning activities.
11. The flipped classroom helps me to develop my problem-solving skills.
12. The flipped classroom facilitates my personalized learning.
13. The flipped classroom is a very enjoyable approach.
14. I prefer the flipped classroom over the traditional lectures.

These 14 items were divided into four dimensions: “content” dimension (items 1-5), such as “the flipped classroom offers me the opportunity to review the lecture as many times as I need to;” “communication” dimension (items 6-8), such as “the flipped classroom helps me to effectively cooperate with my classmates;”
“performance” dimension (items 9-12), such as “the flipped classroom helps me effectively participate in the learning activities;” “interest” dimension (items 13-14), such as “the flipped classroom is a very enjoyable approach.” The Cronbach’s $\alpha$ values of the individual dimensions were 0.85, 0.79, 0.78, and 0.87, respectively, showing acceptable reliability in internal consistency.

Experimental procedure

Figure 3 shows the experimental procedure of this study, which was carried out over a period of 18 weeks in 2017. An experimental group and a control group participated.

- In the experimental group, an online community-based flipped classroom was constructed. There were 33 students in this group, which was taught as follows:
  - In the first two weeks, the students in the experimental group received orientation and preparation for the use of the online community-based flipped learning and participation on the Facebook platform. They also learned basic knowledge of oral skills.
  - Next, for a 5-week period as Time 1, a 5-week period as Time 2, and the last 5-week period as Time 3, the lesson themes were as follows for the experimental group: learning to speak English about topics related to celebrities, music, and food in Time 1, travel, festivals, and animals in Time 2, documentaries, talk shows, science, and others such as speech principles in Time 3. In Time 1, the students learned how to introduce themselves and talk about their interests, such as music or food. And in Time 2, they learned to talk more about life, including trips on holidays, Christmas stories, and animals. Finally, in Time 3, students learned how to express ideas and explain things; therefore, they watched learning videos including documentaries, talk shows, science and other videos. During these 3 time periods and the 9 topics, the teacher would like to enhance students’ English oral performance on the given topics and to promote students’ English expression and oral presentation skills.
  - Out-of-class learning: instructional video learning contents uploaded on Facebook. Video lectures expose students to English listening with scripts before class and encourage them to be responsible for sharing their thoughts or comments on the video using their mobile phone or smartphone. In other words, the
students reviewed the video and said one or two things they had learned from the video and left a comment on Facebook before attending the lecture each week. Students with difficulties doing the out-of-class video learning were directed to individual learning during the in-class time.

- In-class learning: focused on discussion and practice for English speaking as well as writing. Each lesson took up two 50-minute class periods per week; the students worked in pairs or small groups to complete a communicative task in oral or written form. These activities were designed to enhance the students’ English learning, especially their speaking competency. Table 3 provides an illustration and examples of in-class learning activities, including oral English activities such as dialogue practice, short roleplays, and mobile language learning such as roleplay writing and Facebook Live Streaming.

On the other hand, the students in the control group without flipped learning received the same knowledge and information while attending traditional English lectures throughout the semester. The total number of students who participated in this group was 16. In this group, the students were taught using lecture-based and video-based approaches focused on individual discussion activities and pair roleplay in the classroom. Homework was assigned after the class. Students in the control group worked with the same learning activities including oral English activities and mobile language learning, but without Facebook Live Streaming or any other out-of-class learning activities on Facebook.

Table 3. The 18-week in-class learning activities focused on four aspects

<table>
<thead>
<tr>
<th>Oral English activities</th>
<th>Mobile Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>The following activities were used each week with a thematic topic from an oral English textbook. It aims to improve students’ oral fluency practice.</td>
<td>Students used mobile phones to search for information, keywords, download scripts, and send comments on Facebook. They also made video recordings and shared them online.</td>
</tr>
<tr>
<td><strong>Dialogue:</strong> Students role-played a dialogue, guided by a topic and English scripts from the textbook.</td>
<td><strong>Roleplay writing:</strong> Students used mobile devices to gain ideas and information to write their dialogues. They must read and practice to tell the story in class.</td>
</tr>
<tr>
<td><strong>Short Roleplay:</strong> Students role play according to instructions on what they have written in their dialogues from the roleplay writing.</td>
<td><strong>Facebook Live Streaming:</strong> Students recorded dialogues and shared the video clips to get the teacher’s and peers’ immediate feedback.</td>
</tr>
</tbody>
</table>

Results

English oral performance

First, we examined the effectiveness of the proposed approach to ensure that this learning approach could benefit the students’ English oral performance. The inter-rater reliability of the ratings given by the two teachers was 0.764, showing high consistency between their ratings. An independent sample t test was conducted first to compare the three oral performance scores of the two groups.

- **Time 1:** the online community-based flipped classroom scores ($M = 18.60, SD = 1.69$) were higher than the conventional video-based learning classroom scores ($M = 17.12, SD = 1.89$; $t = -2.76, p < .01$).
- **Time 2:** the online community-based flipped classroom scores ($M = 20.36, SD = 1.51$) were higher than the conventional video-based learning classroom scores ($M = 19.15, SD = 1.50$; $t = -2.62, p < .01$).
- **Time 3:** the online community-based flipped classroom scores ($M = 22.12, SD = 1.26$) were higher than the conventional video-based learning classroom scores ($M = 19.62, SD = 1.20$; $t = -6.56, p < .001$).

In addition, one-way analysis of covariance (ANCOVA) was employed to evaluate the students’ oral performance in the experimental group and the control group by adopting the Time 3 scores as the dependent variable and Time 1 scores as the covariate. The test of regression coefficient showed that the assumption of homogeneity for the oral performance scores in Time 1 was not violated ($F = (1, 47) = 0.052, p = .82 > .05$), indicating that ANCOVA could be employed. Table 4 shows the ANCOVA result. It was found that the oral performance of the experimental group was significantly higher than that of the control group in Time 3 ($F = 34.30, p < .001$) by excluding the impact of the scores in Time 1. This result implies that the online community-based flipped learning can more significantly improve students’ learning achievement compared with the conventional video-based instruction. Furthermore, the effect size ($\eta^2$) of flipped learning was 0.43, representing a moderate effect size (Cohen, 1988).
### Table 4. The one-way ANCOVA result of the post test of the two groups

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Adjusted mean</th>
<th>Adjusted SD</th>
<th>F</th>
<th>η²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>33</td>
<td>22.12</td>
<td>1.26</td>
<td>22.10</td>
<td>0.23</td>
<td>34.30***</td>
<td>0.43</td>
</tr>
<tr>
<td>Control</td>
<td>16</td>
<td>19.62</td>
<td>1.20</td>
<td>19.71</td>
<td>0.36</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. ***p < .001.

### Participation between high improvers and low improvers

To further understand the effects of the proposed approach on the learning participation of the students with different degrees of improvement in their learning achievement, the students were classified into high and low improvers based on the ranges of their scores in the final oral test in Time 3 and the first oral test in Time 1. The high improvers (n = 17) improved from 3 to 7 points in this semester, while the low improvers (n = 16) only improved -1 to 2 points. As shown in Figure 4, the high improvers on average participated in the out-of-class learning 4.47 times in Time 1, 3.65 times in Time 2, and 4.18 times in Time 3. In comparison, the low improvers participated on average 4.31 times in Time 1, 3.06 times in Time 2, and 2.75 times in Time 3. In other words, in terms of the high improvers’ and low improvers’ out-of-class learning participation in the flipped speaking on Facebook, the results showed that the high improvers participated much more than the low improvers.

Furthermore, a sample t test was employed to examine the difference in the participation or change for the high and low improvers. The findings are presented in Table 5. The standard errors were 0.94 in Time 1, 1.53 in Time 2 and 1.01 in Time 3 for the high-improver group; and 0.87 in Time 1, 1.18 in Time 2, and 1.29 in Time 3 for the low-improver group. A significant effect was observed in Time 3 (t = 3.54, p < .01). This implies that there was no significant difference or change in the out-of-class learning participation for the high and low improvers in the online community-based flipped classroom in Time 1 or Time 2. However, in Time 3, the high improvers were motivated to engage more and had significantly higher participation compared with the low improvers. Moreover, the effect size (d) of participation for the high improvers was 1.23, representing a moderate effect size (Cohen, 1988).

### Table 5. t-test result of students’ out-of-class learning participation from Time 1 to Time 3 for the high and low improvers

<table>
<thead>
<tr>
<th>18 weeks</th>
<th>High improvers (n = 17)</th>
<th>Low improvers (n = 16)</th>
<th>t</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Time 1</td>
<td>4.47</td>
<td>0.94</td>
<td>4.31</td>
<td>0.87</td>
</tr>
<tr>
<td>Time 2</td>
<td>3.65</td>
<td>1.53</td>
<td>3.06</td>
<td>1.18</td>
</tr>
<tr>
<td>Time 3</td>
<td>4.18</td>
<td>1.01</td>
<td>2.75</td>
<td>1.29</td>
</tr>
</tbody>
</table>

Note. **p < .01.

213
Students’ interactive behaviors

Working from the descriptions of the students’ feedback and responses on Facebook during the 18-week flipped learning, the relationship patterns of the high and low improvers were collected and analyzed. Several previous studies have reported the potential impacts of learning achievement or performance in the flipped classroom, while several researchers have indicated that flipped classrooms could help students engage in tasks or learning activities. Therefore, we further examined the interactive behaviors in the online community flipped classroom by analyzing the feedback patterns of the students with high and low improvement to further investigate their learning analytics. Table 6 shows the frequency and percentage of the individual coded interactive behaviors of the high and low improvers. Among the 15 videos used as the out-of-learning lectures during the semester, 494 interactive behaviors were collected from Facebook. The students’ responses and feedback using mobile technologies were categorized into six types of talk, including clarification (CL) with 214 total occurrences, confirmation (CO) with 97, retelling (RE) with 97, challenge (CH) with 45, suggestions (SU) with 39, and surmise (SR) with only 2. “Clarification” was the most frequent feedback pattern, with an occurrence of 43.0% for the high improvers and 43.8% for the low improvers. Meanwhile, there was 20.4% “confirmation” for the high improvers, 18.6% for the low improvers; 16.5% “retelling” for the high improvers, 23.8% for the low improvers; 10.9% “challenge” for the high improvers, 6.7% for the low improvers; 8.5% “suggestion” for the high improvers, 7.1% for the low improvers and < 5% “surmise” for the high improvers and 0% for the low improvers.

Table 6. The frequency of coded interactive behaviors for the high and low improvers

<table>
<thead>
<tr>
<th>Categories of interactive behaviors</th>
<th>High improvers (n =17)</th>
<th>Low improvers (n =16)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of occurrences</td>
<td>% of occurrences</td>
<td>Number of occurrences</td>
<td>% of occurrences</td>
</tr>
<tr>
<td>CL (Clarification)</td>
<td>122</td>
<td>43.0%</td>
<td>92</td>
</tr>
<tr>
<td>CO (Confirmation)</td>
<td>58</td>
<td>20.4%</td>
<td>39</td>
</tr>
<tr>
<td>RE (Retelling)</td>
<td>47</td>
<td>16.5%</td>
<td>50</td>
</tr>
<tr>
<td>CH (Challenge)</td>
<td>31</td>
<td>10.9%</td>
<td>14</td>
</tr>
<tr>
<td>SU (Suggestion)</td>
<td>24</td>
<td>8.5%</td>
<td>15</td>
</tr>
<tr>
<td>SR (Surmise)</td>
<td>&lt;5</td>
<td>5%</td>
<td>&lt;5</td>
</tr>
<tr>
<td>Ave (Average)</td>
<td>284</td>
<td>100%</td>
<td>210</td>
</tr>
</tbody>
</table>

Figure 5 shows the number of occurrences of the coded interactive behaviors for the high and low improvers. It was found that “clarification,” “confirmation,” “challenge,” “suggestion,” and “surmise” were more commonly used by the students in the high-improvement group than by those in the low-improvement group. Only the number of occurrences of “retelling” was higher in the low-improvement group than in the high-improvement group.

![Figure 5. Number of occurrences of coded interactive behaviors for the high and low improvers](image)

To further examine the six categories of interactive behaviors of the high and low improvers, a sample t test was employed to investigate the significances among the interactive behaviors. According to the results in Table 7, it
was found that the categories of interactive behaviors of clarification (CL) and challenge (CH) for the high improvers were significantly higher than those for the low improvers (CL: \( t = 2.69, p < 0.05, d = 0.94; \) CH: \( t = 2.33, p < 0.05, d = 0.83 \)). This result implies that the students in the high-improvement group exhibited significantly more occurrences of speaking and giving effective responses compared with the low-improvement group.

### Table 7. t-test result of categories of interactive behaviors for the high and low improvers

<table>
<thead>
<tr>
<th>Categories of interactive behaviors</th>
<th>High improvers (n = 17)</th>
<th>Low improvers (n = 16)</th>
<th>( t )</th>
<th>( d )</th>
</tr>
</thead>
<tbody>
<tr>
<td>CL (Clarification)</td>
<td>7.17</td>
<td>1.42</td>
<td>5.75</td>
<td>1.61</td>
</tr>
<tr>
<td>CO (Confirmation)</td>
<td>3.41</td>
<td>2.26</td>
<td>2.43</td>
<td>1.96</td>
</tr>
<tr>
<td>RE (Retelling)</td>
<td>2.76</td>
<td>0.75</td>
<td>3.12</td>
<td>0.34</td>
</tr>
<tr>
<td>CH (Challenge)</td>
<td>1.82</td>
<td>1.38</td>
<td>0.87</td>
<td>0.88</td>
</tr>
<tr>
<td>SU (Suggestion)</td>
<td>1.41</td>
<td>1.46</td>
<td>0.93</td>
<td>0.92</td>
</tr>
<tr>
<td>SR (Surmise)</td>
<td>0.11</td>
<td>0.33</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Ave (Average)</td>
<td>2.78</td>
<td>0.71</td>
<td>2.18</td>
<td>0.36</td>
</tr>
</tbody>
</table>

*Note. \(^* p < .01; ^{**} p < .05.*

### Students’ perceptions of the flipped classroom

The questionnaire survey, modified from Al-Zahrani (2015), was administered and the results analyzed, with the students’ perceptions of the flipped classroom shown in Table 8. The total score revealed high satisfaction with the online community-based flipped learning approach (\( M = 4.06, SD = .23 \)). High satisfaction rates were given for items 1 through 9, and for items 12 and 14. However, items 10, 11, and 13 indicate the students’ moderate satisfaction.

Table 8 shows t-test results of the high and low improvers’ perceptions of the flipped learning. It was found that the high improvers’ perceptions of Communication and Interest had significantly more positive perceptions than those of the low improvers in the dimensions of Communication (\( t = 2.74, p < .05 \)) and Interest (\( t = 2.10, p < .05 \)) with effect sizes 0.96 and 0.72 respectively.

### Table 8. Four dimensions of students’ perceptions of the flipped classroom for the high and low improvers

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>High improvers (n = 17)</th>
<th>Low improvers (n = 16)</th>
<th>( t )</th>
<th>( d )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content (Items 1-5)</td>
<td>4.29</td>
<td>0.69</td>
<td>4.06</td>
<td>0.87</td>
</tr>
<tr>
<td>Communication usefulness (Items 6-8)</td>
<td>4.20</td>
<td>0.60</td>
<td>3.65</td>
<td>0.55</td>
</tr>
<tr>
<td>Performance (Items 9-12)</td>
<td>4.07</td>
<td>0.67</td>
<td>3.63</td>
<td>0.59</td>
</tr>
<tr>
<td>Interest (Items 13-14)</td>
<td>4.62</td>
<td>0.63</td>
<td>4.10</td>
<td>0.80</td>
</tr>
</tbody>
</table>

*Note. \(^* p < .05.*

### Discussion and conclusions

The need to enhance students’ English competency of using technologies and their behavioral patterns has been emphasized (Graf et al., 2009; Hwang, Hsu, Lai, & Hsueh, 2017; Unlu & Wharton, 2015). However, due to the limited learning time and English oral environment, students usually have few opportunities to practice their English oral skills in class. To better understand students’ English oral learning process, oral performance, participation, and interactive patterns, in this study, an online community-based flipped classroom was developed, and a learning analytics approach was employed to explore the learning differences between high improvers and low improvers in the proposed approach. According to the experimental results, the research questions of this study can be answered.

Taken as a whole, the results seem to indicate that the online community-based flipped learning approach using mobile technologies was effective in the English-speaking classroom (Bergmann & Sams, 2015; Sung, 2015). Figure 6 shows the conclusion of this research. First, for the performance aspect, flipped speaking plays a highly positive role in enhancing students’ learning performance over the conventional video-based learning approach. The results revealed that the students seemed to be motivated to learn by accessing a variety of English input or learning materials using their mobile devices. They used mobile technologies to watch videos before the lecture and could complete the pre-class assignments and respond to the comments. In other words, not only does
Flipping speaking help students to learn English better, but it also gives teachers a better understanding of the teacher-student feedback relationship and the students’ learning process (Zhang & Wu, 2016). Additionally, in terms of participation, according to the instructor observations and analysis of the students’ participation during the 18-week course, the students were motivated to participate in the out-of-class learning activities in Time 1 and Time 2, but especially in Time 3, the high improvers had a significantly higher frequency of online participation.

According to the previous studies, Chen, Wang, and Chen, (2014), Han (2015), Jinlei, Ying, and Baohui (2012) have indicated that the flipped language learning demonstrates significant potential for language classroom and enhances learner autonomy. In the flipped classroom, the students had practiced autonomous English learning materials before the lecture and the activities in class. At the end of the semester, the signs of successful development of learner autonomy were also identified, such as their interactive feedback and submissions of the video presentation during Time 1, Time 2 and Time 3. All students submitted at least three videos and responded the interactive behaviors. Moreover, the impact was observed and analyzed for the interactive patterns. The students actively and voluntarily contributed their time and effort to use mobile language learning in and out of the class. Therefore, the study also demonstrated the development of students’ learner autonomy to enhance their language abilities. This helped them quickly engage and learn before the class, encouraging them to become more responsible people (Cummins, 2016; Han, 2015; Sung, 2015).

The online flipped classroom also helps teachers to manage the classroom time more efficiently (Cockrum, 2013). Third, according to the data from the students’ interactions using the proposed approach, we explored factors which may influence the nature of the interactive behaviors in the online flipped speaking classroom. Both high and low improvers showed awareness of “Clarification,” “Challenge,” “Retelling,” “Surmise,” Confirmation,” and “Suggestion” in the speaking classroom. The results showed that the high improvers exhibited a significantly higher frequency of “Clarification” and “Challenge,” suggesting that teachers could help low improvers possibly desire them, but this does not necessarily mean that students in flipped speaking classrooms equally desire these interactive behaviors. Teachers could use ways to stimulate students to communicate and to use different interactive behaviors to express ideas. The online flipped classroom may be an effective learning approach to provide students with opportunities to reflect on their practices in discussion with peers and teachers. Finally, for the perception aspect, the research findings suggest that the high improvers were significantly more satisfied with the online community-based flipped classroom than low improvers. As the high improvers also expressed more positive percpetions of communication and interest, teachers could consider leading in additional strategies to raise the interest and communication of low improvers in the future.
To sum up, the online community-based flipped learning approach using mobile technologies offers a rich, informal, and ubiquitous learning environment in which it is possible for students and teachers to better control English language teaching and learning and to improve the learners’ language proficiency (Hsieh et al., 2017). This study explored students’ oral performance, participation, interactive patterns, and their perceptions of the flipped classroom when using the proposed approach, and the findings highlighted how the online community-based flipped approach improved the students’ learning attitudes and dealt with the issues of student engagement with feedback using mobile technologies for flipping speaking on Facebook. With these results, regarding the students’ oral performance, participation, and interactive behavior, classroom lecturers using the proposed approach can consider more benefits over the conventional video-based English speaking classroom. Flipping speaking could be successfully recognized as an effective learning approach in this study. This proposed approach not only helps teachers reach students of different abilities or learning achievement, but also improves classroom management, giving teachers more time to interact with each student (Basal, 2015; Bergmann & Sams, 2015; Sung, 2015). Moreover, the flipped speaking classroom helps students become more responsible for their learning (DeLozier & Rhodes, 2017; Han, 2015; Sung, 2015), and allows teachers more individual interaction with students via the online community platform and the ability to develop better relationships with their students (Zhang & Wu, 2016).

In the future research on flipped English speaking classrooms, teachers could use more strategies to help students learn the concept of adapting their use of language to conform to standards or traditions in the given contexts, such as giving advice, asking questions, or proposing different ideas. In addition, by using the online community-based flipped learning and teaching, teachers could enhance students’ critical thinking and the collaborative relationship between the students and teacher (Unlu & Wharton, 2015). Consequently, several studies can be considered in the future, such as an investigation of the effects of interactive behaviors comparing different levels of autonomous learners or students’ learning performance with different flipped classroom approaches. Moreover, further research can probe the teacher-student use of the mobile-flipped pedagogical approach, such as planning, reflecting, or the detailed out-of-class learning and in-class learning activities using appropriate learning content so as to promote low improvers’ English oral performance, participation, and perceptions of the flipped classroom for better learning. Finally, further investigations into flipping speaking in English can provide a more holistic picture of English language teaching development. In the future, it would be worth investigating the impacts of the approach on students’ English learning performance and perceptions in other dimensions. Moreover, it could be valuable to investigate the impacts of the approach on the learning performances of the students with different personal characteristics.

References


Bergmann, J., & Sams, A. (2012). Flip your Classroom: Reach every student in every class every day. Alexandria, VA: International Society for Technology in Education.


Applying Learning Analytics for the Early Prediction of Students’ Academic Performance in Blended Learning

Owen H. T. Lu¹, Anna Y. Q. Huang¹, Jeff C. H. Huang², Albert J. Q. Lin¹, Hiroaki Ogata³ and Stephen J. H. Yang¹²

¹Department of Computer Science and Information Engineering, National Central University, Taiwan // ²Department of Computer Science and Information Engineering, Hwa Hsia University of Technology, Taiwan // ³Graduate School of Informatics, Kyoto University, Japan

ABSTRACT

Blended learning combines online digital resources with traditional classroom activities and enables students to attain higher learning performance through well-defined interactive strategies involving online and traditional learning activities. Learning analytics is a conceptual framework and as a part of our precision education used to analyze and predict students’ performance and provide timely interventions based on student learning profiles. This study applied learning analytics and educational big data approaches for the early prediction of students’ final academic performance in a blended Calculus course. Real data with 21 variables were collected from the proposed course, consisting of video-viewing behaviors, out-of-class practice behaviors, homework and quiz scores, and after-school tutoring. This study applied principal component regression to predict students’ final academic performance. The experimental results show that students’ final academic performance could be predicted when only one-third of the semester had elapsed. In addition, we identified seven critical factors that affect students’ academic performance, consisting of four online factors and three traditional factors. The results showed that the blended data set combining online and traditional critical factors had the highest predictive performance.

Keywords

Learning analytics, Educational big data, MOOCs, Blended learning, Principal component regression

Introduction

Blended learning, also known as hybrid learning or mixed-mode instruction, incorporates one or two learning strategies into traditional classroom teaching. In 1960, many computer programming courses relied on the Internet to deliver digital learning materials to students; for example, Programmed Logic for Automatic Teaching Operations, developed at the University of Illinois (Hart, 1995), provided teaching activities that could be conducted on a large scale to enable a single instructor to simultaneously teach a large number of students.

In recent years, blended learning has become a popular teaching strategy because of the development of data analysis and computation; for example, Ellis, Pardo, and Han (2016) integrated social networking into a one-semester course and monitored the behaviors of over 220 undergraduate engineering students. The researchers used the students’ interactive records to examine how to help them succeed in a collaboratively driven course. Hong et al. (2016) adopted a web game to develop ten teaching scenarios. After 6 weeks of experimentation on 110 elementary school students, the researchers indicated that the students were highly motivated by the combination of game-based learning and traditional classroom activities. Huang, Yang, Chiang, and Su (2016) improved students’ learning motivations and performance in an English course by incorporating a mobile-based vocabulary feedback application into a traditional classroom environment.

To gain benefits from blended learning, many educators have adopted the Online Assessment System (OAS) or Massive Open Online Courses (MOOCs) into their course design; for example, Awang and Zakaria (2013) integrated the OAS into an integral course for 101 college students. The results indicated that the OAS improved the students’ learning performance. Lu, Huang, Huang, and Yang (2017) incorporated MOOCs into a course and the results showed evidence of a well-defined intervention strategy. The course not only facilitated the students’ learning achievements but also increased their level of engagement. Although the aforementioned studies have explained the advantages of blended learning, many researchers have asserted that in blended courses, monitoring students’ learning behaviors and habits is difficult because of the complex learning environment (Ellis et al., 2016; Hong et al., 2016; Huang et al., 2016). Furthermore, at-risk students cannot be identified, and thus timely interventions cannot be conducted to facilitate learning success (Tempelaar, Rinties, & Giesbers, 2015).
To help students achieve classroom success, educators in Europe and the United States have recently applied learning analytics. In 2011, Horizon Report, a report of educational trends, investigated the benefits and future trends of learning analytics (Johnson, Smith, Willis, Levine, & Haywood, 2011). The report defined learning analytics as an ideal framework to improve learning performance based on data of students’ learning history, because of the limitations of data analysis and computation, learning analytics has been considered as a conceptual framework since 2011. Because of the rise of big data technology, in 2016, a special issue of Horizon Report was released on learning analytics to highlight that the optimal time to incorporate learning analytics into classroom settings had arrived (Johnson et al., 2016).

In recent years, learning analytics has served as a conceptual framework for the analysis of course characteristics, and has included prediction of students’ learning performance, educational data analysis process development (Hwang, Chu, & Yin, 2017), data collection, and timely intervention (Hwang, 2014). To develop a conceptual framework for learning analytics, many researchers have designed and implemented courses with strategies for learning analytics. Lu et al. (2017) measured student engagement in a virtual learning environment and intervened with the students’ learning activities according to the engagement score. The results showed improvements in the students’ final academic performance and their self-regulated abilities after applying learning analytics. Hachey, Wladis, and Conway (2014) collected the learning data of 962 students to determine the factors that influence their grade point averages (GPAs). The results showed that students with no experience of online learning obtained low retention rates and had low GPA scores. The researchers concluded that online learning and practice must be offered to students without relevant experience before the beginning of a course (Papamitsiou & Economides, 2014).

In our research, learning analytics is a conceptual framework and as a part of our Precision education used to analyze and predict students’ performance and provide timely interventions based on student learning profiles. The idea of our Precision education is the same as of The Precision Medicine Initiative (see https://obamawhitehouse.archives.gov/node/333101), which was proposed by President Obama in his 2015 State of the Union address, the Initiative is a new research effort to revolutionize the medical treatment of disease. As addressed in this Initiative, most treatments were designed for the average patients as a result of one-size-fits-all approach treatments which could be successful for some patients but not for others. With the same philosophy, we carry the idea of Precision medicine, which is to improve the diagnosis, prediction, treatment, and prevention of disease, and define the objective of our Precision education as the improvement of diagnosis, prediction, treatment, and prevention of learning outcome.

The previous studies have shown that the development of big data technology has enabled learning analytics to become a suitable method for facilitating student success. The advantage of blended learning is that huge quantities of learning data can be collected through learning management system (LMS) to enrich personal learning data. However, few case studies have been conducted on the effects of applying learning analytics in blended courses due to the complexity of learning environments and the diversity of data. To provide timely interventions for at-risk students through learning analytics in blended learning, the present study not only implemented a MOOC and OAS enabled Calculus course but also proposed a process for the early identification of at-risk students. To predict students’ final academic performance, many studies have used only one data set: a subset of a blended course. To improve prediction performance, critical factors may need to be identified and prediction accuracy may need to be compared using a data set combining online and traditional learning activities. The following research questions were proposed:

**RQ1.** How early can we predict students’ final academic performance?
**RQ2.** Which are the most critical factors that affect students’ final academic performance in blended learning?
**RQ3.** Which type of data set (blended vs. online vs. traditional) is more effective for predicting students’ final academic performance in blended learning?

**Literature review**

**Identification of at-risk students**

According to the learning analytics executive reports by Arroway, Morgan, O’Keefe, and Yanosky (2015) and Kuzilek, Hlosta, Herrmannova, Zdrahal, and Wolff (2015), the first stage of implementing learning analytics is to identify at-risk students. Moreover, at-risk student identification must be conducted as early as possible to allow sufficient time for instructors to conduct educational interventions to facilitate students’ learning achievements. Early at-risk student identification originated from the implementation of an open course that yielded a high dropout rate (Yang, Huang, & Huang, 2017).
Many researchers have defined dropout as a risk of MOOCs and have designed prediction methods to identify the dropout group. Xing, Chen, Stein, and Marcinkowski (2016) collected data on 3,617 students’ video watching behaviors in 2014 and developed a classification model to identify the students likely to drop out by the following week. The results suggested that the retention rate would have been higher if the instructors had conducted timely interventions based on the prediction results. Lara, Lizcano, Martínez, Pazos, and Riera (2014) collected historical data on 100 students in a virtual learning environment consisting of five variables and proposed a knowledge discovery system for dividing students into dropout and non-dropout groups. The researchers reached a 90% classification accuracy through a verification process involving 100 students. Thammasiri, Delen, Meesad, and Kasap (2014) compared several resample algorithms with 7 years of student interaction data to assess data imbalance. Moreover, the target data was 80% true, indicating that 80% of freshman continued their studies, and 20% as false, indicating that 20% dropped out. These results show that the combination of synthetic minority oversampling (SMOTE) and the support vector machine yielded a classification accuracy of 90%, which was an improvement on the 86% accuracy without resampling in 10-fold cross validation. In addition to online courses, numerous researchers have incorporated student learning performance prediction into traditional classroom settings. Hachey et al. (2014) used a unique combination of variables to construct several classification models and verified the models with historical data collected from a learning management system. The results indicated that if the goal is to predict the learning outcomes of students with online course experience, retention rate is a more useful variable than GPA. For all other goals, GPA is more favorable. The results of the aforementioned studies show that at-risk students can be identified through classification methods if at-risk is defined as potential course dropout. However, in contrast to some studies, which have used data from open courses and pure online courses, another group of researchers defined at-risk as students who failed or obtained low grades at the end of a course. Many researchers have since adopted this approach for predicting students’ final academic performance.

**Students’ final academic performance prediction**

To identify at-risk students based on their final grades, scores, or learning outcomes, educational data mining can be used to identify students’ behavioral patterns and predict their grades (Romero & Ventura, 2010). Romero, López, Luna, and Ventura (2013) collected data on 114 students from an online discussion forum and separated them into several data subsets on a weekly basis before evaluating each data set’s predictive accuracy through several data-mining methods. Romero et al. (2013) used the sequential minimal optimization classification algorithm and student interaction data before a midterm exam to achieve the highest accuracy for predicting student learning performance. Hu, Lo, and Shih (2014) developed an early warning system by using a decision tree classifier. The model was constructed from data on 300 students and contained 13 online variables, including for how long each student had used the system and how many documents had been read by each student in the preceding week. The results revealed a 95% accuracy in predicting whether students would pass or fail based on 1–4 weeks of data from a skewed data set. To verify which critical factors affect prediction performance, Villagra -Arnedo, Gallego-Durán, Compañ, Llorens-Largo, and Molina-Carmona (2016) determined 8 variables for student behavior and 53 for learning activity from a learning management system. Villagrá-Arnedo et al. (2016) designed four experiments to validate a data set with different variable combinations. The results demonstrated that a data set with particular variables had the highest correlation coefficient with grades and could attain higher prediction accuracy than the others.

In addition to predicting student learning outcomes, one study used students’ grades as prediction labels and marked students as at-risk if their prediction grades were below average. Meier, Xu, Atan, and van der Schaar (2016) used regression to design a neighbourhood selection process to predict students’ grades. The researchers claimed that the proposed algorithm achieved 76% accuracy. Asif, Merceron, and Pathan (2014) used a naive Bayes classifier to demonstrate that students’ grades in their final year of university could be predicted based on student data collected during freshman year. In addition, the researchers executed the feature selection process before classification and the results showed that the data set from which socioeconomic and demographic variables had been removed was reasonably accurate. Huang and Fang (2013) used students’ final grades as prediction targets. To evaluate the prediction results, the researchers designed two quantitative indicators to transfer the regression mean square error into prediction accuracy. The final results showed that the students’ final exam scores were predictable to 88% accuracy based on eight variables collected from a learning management system. Previous studies have explained that “at-risk” can generally be used to describe students who dropout, fail, or achieve low grades on courses. We can fulfill the critical requirement of learning analytics by using students’ final grades or scores as prediction indicators and designing a data-mining methodology based on classification or regression for the early prediction of indicators.
Recent studies have used data collected from entire course periods, which is problematic because, through this method, students can only be determined as at-risk after the conclusion of a course, which is ineffective in real scenarios. Moreover, recent studies have used single data sets collected from virtual learning environments or classroom activities, which is ineffective for applying the results to blended courses that combine online and face-to-face learning. Therefore, we referred to recent studies to define the following four aspects for consideration: First, data must be divided into sub data sets based on duration (Hu et al., 2014; Romero et al., 2013). Second, critical factors must be identified to improve prediction accuracy (Asif et al., 2014; Villagrá-Arnedo et al., 2016); for example, Villagrá-Arnedo et al. (2016) reduced the number of variables from 61 to 23 without losing prediction accuracy. Third, a predesigned regression model used in previous studies called principle component regression (PCR) (Agudo-Peregrina, Iglesias-Pradas, Conde-González, & Hernández-García, 2014; Çevik, 2015; Huang & Fang, 2013; Meier et al., 2016) was used. The model was also implemented and evaluated in our previous study. PCR involves performing principle component analysis (PCA) to calculate the principle components, some of which can be used as variables in multiple linear regression. Fourth, design indicators and acceptance criteria must be considered to evaluate prediction performance. Although the regression model provided several indicators to evaluate performance, it did not provide any accuracy indicator. Therefore, following the concept of prediction accuracy proposed by Huang and Fang (2013), we applied the cross-validation mechanism proposed by Golub, Heath, and Wahba (1979) to design indicators to evaluate prediction performance. Moreover, in recent studies, the acceptance of prediction accuracy ranged from 75% (Villagrá-Arnedo et al., 2016) to 95% (Hu et al., 2014).

Method and experiments

Participation and learning activities

The participants in this study were 33 male and 26 female students. The experiment was conducted in a Calculus course that ran from September 2015 to February 2016. This study utilized MOOCs and the OAS to improve freshman students’ learning outcomes at a university in Northern Taiwan.

![Calculus course learning activities](image)

**Figure 1. Calculus course learning activities**

<table>
<thead>
<tr>
<th>Weeks</th>
<th>Homework</th>
<th>Quiz</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>H1</td>
<td>Q1</td>
</tr>
<tr>
<td>2</td>
<td>H2</td>
<td>Q2</td>
</tr>
<tr>
<td>3</td>
<td>H3</td>
<td>Q3</td>
</tr>
<tr>
<td>4</td>
<td>H4</td>
<td>Q4</td>
</tr>
<tr>
<td>5</td>
<td>H5</td>
<td>Q5</td>
</tr>
<tr>
<td>6</td>
<td>H6</td>
<td>Q6</td>
</tr>
<tr>
<td>7</td>
<td>H7</td>
<td>Q7</td>
</tr>
<tr>
<td>8</td>
<td>H8</td>
<td>Q8</td>
</tr>
<tr>
<td>9</td>
<td>H9</td>
<td>Q9</td>
</tr>
</tbody>
</table>

**Table 1. Homework and quiz execution weeks**

<table>
<thead>
<tr>
<th>Week</th>
<th>Content</th>
<th>Week</th>
<th>Content</th>
<th>Week</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Function Limitation</td>
<td>7</td>
<td>Anti-differentiation</td>
<td>13</td>
<td>Vector Space</td>
</tr>
<tr>
<td>2</td>
<td>Differentiation</td>
<td>8</td>
<td>Number Integral</td>
<td>14</td>
<td>Curve in Space</td>
</tr>
<tr>
<td>3</td>
<td>Newton's Method</td>
<td>9</td>
<td>Harmonic series</td>
<td>15</td>
<td>Surface</td>
</tr>
<tr>
<td>4</td>
<td>Integral</td>
<td>10</td>
<td>Taylor Error</td>
<td>16</td>
<td>Scalar Field</td>
</tr>
<tr>
<td>5</td>
<td>Piecewise Function</td>
<td>11</td>
<td>Fourier Series</td>
<td>17</td>
<td>Multiple Integral</td>
</tr>
<tr>
<td>6</td>
<td>Arc Length</td>
<td>12</td>
<td>Polar</td>
<td>18</td>
<td>Line Integral</td>
</tr>
</tbody>
</table>

**Table 2. Course content presented over 18 weeks (see http://mathweb.math.ncu.edu.tw/calc/maple-tutorial.html)**

The Calculus course lasted for 18 weeks and included six learning activities (Figure 1). During the course, the participants used MOOCs to preview Calculus content through Open edX (see https://open.edx.org/about-open-edx) and practiced Calculus by using the OAS through Maple T.A. (see http://www.maplesoft.com/). To improve
participants’ mathematics ability, an instructor provided weekly after-school tutoring for each participant. To encourage the participants to continue studying Calculus, the instructor assigned paper homework exercises. To evaluate the students’ learning performance for each topic, the instructor administered quizzes for specific weeks. The weekly quizzes, homework assignments, and course content are listed in Table 1 and Table 2.

Data sets of learning activities and variables

The MOOC and OAS enabled Calculus course collected participant learning profiles, which consisted of their video-viewing behaviors, out-of-class practice, homework assignments, and quiz scores. In particular, this study collected data on video-viewing behaviors from Open edX and data on out-of-class practice from the Maple T.A. Both types of data were categorized as online behavior. Table 3 lists the data variables definition for the Calculus course.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Category</th>
<th>Learning environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>Number of days a student exhibits activity* per week</td>
<td>Online</td>
<td>MOOCs</td>
</tr>
<tr>
<td>X2</td>
<td>Number of activities’ a student engages in per week</td>
<td>Online</td>
<td></td>
</tr>
<tr>
<td>X3</td>
<td>Number of days a student watches videos per week</td>
<td>Online</td>
<td></td>
</tr>
<tr>
<td>X4</td>
<td>Number of videos a student watches per week**</td>
<td>Online</td>
<td></td>
</tr>
<tr>
<td>X5</td>
<td>Number of videos a student completely watches*** per week</td>
<td>Online</td>
<td></td>
</tr>
<tr>
<td>X6</td>
<td>Number of times a student clicks “Forward seek” or “Backward seek” during video viewing per week</td>
<td>Online</td>
<td></td>
</tr>
<tr>
<td>X7</td>
<td>Number of videos during which a student clicks “Pause” per week</td>
<td>Online</td>
<td></td>
</tr>
<tr>
<td>X8</td>
<td>Number of videos during which a student clicks “Stop” per week</td>
<td>Online</td>
<td></td>
</tr>
<tr>
<td>X9</td>
<td>Number of times a student clicks “Play” per week</td>
<td>Online</td>
<td></td>
</tr>
<tr>
<td>X10</td>
<td>Number of times a student clicks “Forward seek” per week</td>
<td>Online</td>
<td></td>
</tr>
<tr>
<td>X11</td>
<td>Number of times a student clicks “Backward seek” per week</td>
<td>Online</td>
<td></td>
</tr>
<tr>
<td>X12</td>
<td>Number of times a student clicks “Pause” per week</td>
<td>Online</td>
<td></td>
</tr>
<tr>
<td>X13</td>
<td>Number of times a student clicks “Stop” per week</td>
<td>Online</td>
<td></td>
</tr>
<tr>
<td>X14</td>
<td>Number of times a student engages in online practice per week</td>
<td>Online</td>
<td>OAS</td>
</tr>
<tr>
<td>X15</td>
<td>Number of Calculus units a student practices per week</td>
<td>Online</td>
<td></td>
</tr>
<tr>
<td>X16</td>
<td>Number of days a student engages in online practice per week</td>
<td>Online</td>
<td></td>
</tr>
<tr>
<td>X17</td>
<td>Sum of days of practiced Calculus units per week</td>
<td>Online</td>
<td></td>
</tr>
<tr>
<td>X18</td>
<td>Student’s weekly practice score</td>
<td>Online</td>
<td></td>
</tr>
<tr>
<td>X19</td>
<td>Student’s weekly homework score</td>
<td>Traditional</td>
<td>Paper</td>
</tr>
<tr>
<td>X20</td>
<td>Student’s weekly quiz score</td>
<td>Paper</td>
<td></td>
</tr>
<tr>
<td>X21</td>
<td>Number of times a student participates in after-school tutoring per week</td>
<td>Classroom</td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>Student’s final academic performance</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *MOOC activity refers to logging in to watch videos or browse course content. **Counting only once if repeated; unfinished video viewing is included. ***Completely” refers to more than 95%.

Process for predicting students’ final academic performance

At-risk students can be identified as those with a predicted final academic performance of lower than 60. In the blended Calculus course, we applied a final academic performance prediction process with PCR consisting of data preprocessing, modeling, and evaluation phases. The data preprocessing phase consisted of data integration and data set separation. Data integration focused on integrating the learning data derived from MOOCs, the OAS, homework, quiz scores, and after-school tutoring. This study defined 21 variables from the blended learning environments consisting of data of online and traditional learning. The details of variables are described in Table 3. In the data set separation, the duration of the collected learning data was identified. The details of the proposed accumulated and duration data sets are described in the following section. In the modeling phase, a prediction model for students’ final academic performance was generated through PCR. The evaluation phase was focused on measuring the goodness of fit and predictive effectiveness of the regression model. In the evaluation phase, this study measured not only the goodness of fit of the regression model by using the mean squared error (MSE), coefficient of determination (R^2), and Quantile–Quantile (Q–Q) plot but also the predictive performance of the regression model by using the predictive MSE (pMSE) and predictive mean absolute percentage correction (pMAPC), both of which were proposed in our previous study.
Experimental data set description

To investigate the influence of data set duration on predictive effectiveness, this study proposed accumulated and duration data sets. The purpose of the accumulated data set was to record learning data collected from the first week to a specified week, whereas that of the duration data set was to record the participants’ learning behaviors during specific weeks. \( W_i^1 \) indicates that the data set has collected data on the participants’ learning behaviors from week \( i \) to week \( j \). The accumulated and duration data sets included \( W_i^6, W_i^{12} \) and \( W_i^{13} \) data sets and \( W_i^{14} \) data set, respectively. \( W_i^6, W_i^{12} \), and \( W_i^{14} \) were the three accumulated data sets that recorded students’ learning behaviors from weeks 1-6, 1-12, and 1-18, respectively. \( W_i^{13} \) and \( W_i^{15} \) were the two duration data sets that recorded students’ learning behaviors from weeks 7-12 and 13-18, respectively. The statistics for variables \( X_1-X_{21} \) based on the accumulated (\( W_i^6, W_i^{12}, \) and \( W_i^{14} \)) and duration (\( W_i^{13} \) and \( W_i^{15} \)) data sets are listed in Table 4 and Error! Reference source not found..

<table>
<thead>
<tr>
<th>Variable</th>
<th>Data set ( W_i^1 )</th>
<th>Data set ( W_i^{12} )</th>
<th>Data set ( W_i^{13} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Scale</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>( X_1 )</td>
<td>0.0-4.17</td>
<td>2.33</td>
<td>0.96</td>
</tr>
<tr>
<td>( X_2 )</td>
<td>0.0-1410.33</td>
<td>482</td>
<td>254.34</td>
</tr>
<tr>
<td>( X_3 )</td>
<td>0.0-3.0</td>
<td>1.26</td>
<td>0.66</td>
</tr>
<tr>
<td>( X_4 )</td>
<td>0.0-10.33</td>
<td>4.26</td>
<td>2.67</td>
</tr>
<tr>
<td>( X_5 )</td>
<td>0.0-10.0</td>
<td>2.7</td>
<td>2.3</td>
</tr>
<tr>
<td>( X_6 )</td>
<td>0.0-7.33</td>
<td>2.42</td>
<td>1.86</td>
</tr>
<tr>
<td>( X_7 )</td>
<td>0.0-7.83</td>
<td>3.07</td>
<td>2.05</td>
</tr>
<tr>
<td>( X_8 )</td>
<td>0.0-9.67</td>
<td>2.37</td>
<td>2.21</td>
</tr>
<tr>
<td>( X_9 )</td>
<td>0.0-309.33</td>
<td>48.96</td>
<td>55.58</td>
</tr>
<tr>
<td>( X_{10} )</td>
<td>0.0-154.83</td>
<td>13.99</td>
<td>23.36</td>
</tr>
<tr>
<td>( X_{11} )</td>
<td>0.0-28.5</td>
<td>4.92</td>
<td>5.71</td>
</tr>
<tr>
<td>( X_{12} )</td>
<td>0.0-43.5</td>
<td>11.47</td>
<td>10.34</td>
</tr>
<tr>
<td>( X_{13} )</td>
<td>0.0-11.5</td>
<td>2.61</td>
<td>2.5</td>
</tr>
<tr>
<td>( X_{14} )</td>
<td>0.0-8.5</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>( X_{15} )</td>
<td>0.0-2.17</td>
<td>1.55</td>
<td>0.62</td>
</tr>
<tr>
<td>( X_{16} )</td>
<td>0.0-2.33</td>
<td>1.09</td>
<td>0.51</td>
</tr>
<tr>
<td>( X_{17} )</td>
<td>0.0-3.17</td>
<td>1.8</td>
<td>0.79</td>
</tr>
<tr>
<td>( X_{18} )</td>
<td>0.0-9.12</td>
<td>5.99</td>
<td>2.33</td>
</tr>
<tr>
<td>( X_{19} )</td>
<td>0.0-9.99</td>
<td>9.09</td>
<td>1.61</td>
</tr>
<tr>
<td>( X_{20} )</td>
<td>0.0-9.94</td>
<td>7.83</td>
<td>1.85</td>
</tr>
<tr>
<td>( X_{21} )</td>
<td>0.0-4.0</td>
<td>0.14</td>
<td>0.6</td>
</tr>
</tbody>
</table>

In Table 4 and Error! Reference source not found., “Scale” denotes the variable range from the minimum to maximum value. “Mean” and “SD” indicate the average and standard deviation values of 59 students, respectively. In the Calculus course, the average and standard deviation of the participants’ scores were 70.05 and 19.2, respectively. The minimum and maximum Calculus scores were 25 and 100, respectively.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Data set ( W_i^{12} )</th>
<th>Data set ( W_i^{13} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Scale</td>
<td>Mean</td>
</tr>
<tr>
<td>( X_1 )</td>
<td>0.0-3.33</td>
<td>1.38</td>
</tr>
<tr>
<td>( X_2 )</td>
<td>0.0-537.33</td>
<td>161.21</td>
</tr>
<tr>
<td>( X_3 )</td>
<td>0.0-2.5</td>
<td>0.82</td>
</tr>
<tr>
<td>( X_4 )</td>
<td>0.0-10.5</td>
<td>3.21</td>
</tr>
<tr>
<td>( X_5 )</td>
<td>0.0-8.83</td>
<td>1.95</td>
</tr>
<tr>
<td>( X_6 )</td>
<td>0.0-7.83</td>
<td>1.79</td>
</tr>
<tr>
<td>( X_7 )</td>
<td>0.0-7.67</td>
<td>2.32</td>
</tr>
<tr>
<td>( X_8 )</td>
<td>0.0-8.17</td>
<td>1.74</td>
</tr>
<tr>
<td>( X_9 )</td>
<td>0.0-247.33</td>
<td>37.87</td>
</tr>
<tr>
<td>( X_{10} )</td>
<td>0.0-68.83</td>
<td>7.68</td>
</tr>
<tr>
<td>( X_{11} )</td>
<td>0.0-30.33</td>
<td>3.6</td>
</tr>
<tr>
<td>( X_{12} )</td>
<td>0.0-32.67</td>
<td>7.28</td>
</tr>
<tr>
<td>( X_{13} )</td>
<td>0.0-9.0</td>
<td>1.89</td>
</tr>
</tbody>
</table>
Regression model estimation

The performance indicators for evaluating the prediction results in this study were the pMSE and pMAPC, both of which were proposed in our previous study. In the present study, we introduced 10-fold cross validation with shuffling to calculate the pMSE and pMAPC values. We used the testing data obtained from the 10-fold cross validation to calculate the prediction performance. The pMSE and pMAPC equations are as follows:

\[ pMSE = \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} (p_i - a_i)^2, \quad p_i \in p^{test} \text{ and } a_i \in A \]  

\[ pMAPC = 1 - \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} \left| \frac{p_i - a_i}{\sigma} \right|, \quad p_i \in p^{test} \text{ and } a_i \in A \]  

The symbols \( a_i \) and \( p_i \) represent the actual and predictive scores, respectively, of student \( s_i \). \( A = \{ a_1, a_2, ..., a_{n_{test}} \} \) records each student’s Calculus score. The symbol \( \bar{a} \) represents the average score of all students in the blended Calculus course. \( p^{test} = [p_1, p_2, ..., p_{n_{test}}] \) records the predictive Calculus score in the testing data. A lower pMSE value and higher pMAPC value indicate higher predictive performance and higher predictive accuracy, respectively. Therefore, our objective was to find a regression model with a lower pMSE and higher pMAPC.

Experimental results and discussion

Earliness of students’ final academic performance prediction

Regression Model Estimation

We applied PCR to five data sets and generated 21 final academic performance prediction models for each data set. Table 5 lists the average values and scale of the \( R^2 \), adjusted \( R^2 \), and Durbin-Watson statistic for each data set. The Durbin-Watson values indicate that the 21 learning variables are independent. The ranges of the average \( R^2 \) and adjusted \( R^2 \) values for each data set are 0.34-0.47 and 0.30-0.38, respectively. These results are similar to those of previous studies (Agudo-Peregrina et al., 2014; Çevik, 2015), which indicates that the explanatory power of each regression model in the present study was acceptable. Regarding the scale of the \( R^2 \) and adjusted \( R^2 \), the scale ranges of the accumulated data sets are all higher than the scales of the duration data sets, which suggests that the explanatory power of the regression models using the accumulated data sets was higher than that of the regression models using the duration data sets.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Mean ( R^2 )</th>
<th>Scale</th>
<th>Mean Adjusted ( R^2 )</th>
<th>Scale</th>
<th>Mean Durbin-Watson</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accumulated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>data set</td>
<td>( W_8^1 )</td>
<td>0.47</td>
<td>( W_{12}^1 )</td>
<td>0.16-0.66</td>
<td>0.37</td>
<td>0.15-0.52</td>
</tr>
<tr>
<td></td>
<td>( W_8^2 )</td>
<td>0.47</td>
<td>( W_{12}^2 )</td>
<td>0.11-0.69</td>
<td>0.36</td>
<td>0.08-0.52</td>
</tr>
<tr>
<td></td>
<td>( W_8^3 )</td>
<td>0.48</td>
<td>( W_{12}^3 )</td>
<td>0.10-0.72</td>
<td>0.38</td>
<td>0.08-0.56</td>
</tr>
<tr>
<td>Duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>data set</td>
<td>( W_{12}^4 )</td>
<td>0.34</td>
<td>( W_{12}^5 )</td>
<td>0.01-0.70</td>
<td>0.31</td>
<td>0.02-0.53</td>
</tr>
<tr>
<td></td>
<td>( W_{12}^6 )</td>
<td>0.43</td>
<td>( W_{12}^7 )</td>
<td>0.03-0.59</td>
<td>0.30</td>
<td>0.01-0.43</td>
</tr>
</tbody>
</table>

Regarding testing of the regression models, Table 6 lists the values of the F-test and corresponding significance level for each data set. Datasets \( W_{14}^1 \), \( W_{14}^2 \), \( W_{14}^3 \), and \( W_{14}^4 \) had 21, 20, 16, and 17 regression models, respectively. According to the conventional estimation results in Table 5 and Table 6, the accumulated data sets had regression models with better goodness of fit than those of the duration data sets.

226
Table 6. F-test values and corresponding significance levels for five data sets

<table>
<thead>
<tr>
<th>Data set</th>
<th>Value of F-test</th>
<th>p-value of F-test</th>
<th>Number of significant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Scale</td>
<td>Mean</td>
</tr>
<tr>
<td>Accumulated data set</td>
<td>4.93</td>
<td>3.29~11.24</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>4.50</td>
<td>2.32~7.25</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>4.75</td>
<td>2.21~6.53</td>
<td>0.007</td>
</tr>
<tr>
<td>Duration data set</td>
<td>3.43</td>
<td>0.55~5.31</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>3.43</td>
<td>0.72~5.90</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Table 7 lists the prediction indicators for the five data sets. The pMSE and pMAPC ranges among the data sets are 214-248 and 0.82-0.83, respectively. Regarding the mean of the pMSE, the accumulated data sets had slightly lower means than did the duration data sets. However, according to the pMSE values, the predictive error for each participant’s final academic performance in each of the five data sets was close to 15. By contrast, the mean range of the pMAPC among the accumulated and duration data sets was 0.82-0.83. Regarding the average pMSE and pMAPC values, predictive performance was fairly similar in the accumulated and duration data sets because some information may have been lost when computing the average. To solve this problem, this study conducted Wilcoxon signed-rank testing for the 21 regression models for each data set.

The results of Wilcoxon signed-rank testing of the five data sets are listed in Table 7. The Wilcoxon signed-rank test results for pMSE and pMAPC values are listed in the lower and upper triangular matrices, respectively. For the Wilcoxon signed-rank tests for pMSE and pMAPC, the accumulated data sets $W_6^5$ and $W_6^8$ had significantly different results to the duration data sets $W_6^{12}$ and $W_6^{13}$, suggesting that the predictive performance was significantly different between the data set types. Furthermore, we applied box plots to determine which accumulated data set had the highest predictive performance.

Table 7. Results of predictive performance for the five data sets

<table>
<thead>
<tr>
<th></th>
<th>Mean of pMSE</th>
<th>Mean of pMAPC</th>
<th>pMSE (Wilcoxon signed-rank test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accumulate data set</td>
<td>214.85</td>
<td>0.82</td>
<td>$W_6^5$ 0.00** $W_6^{12}$ 0.01* $W_6^{10}$ 0.04* $W_6^{12}$ 0.00** $W_6^{13}$ 0.00***</td>
</tr>
<tr>
<td></td>
<td>230.70</td>
<td>0.82</td>
<td>0.54 0.00** 0.03* 0.07 0.04*</td>
</tr>
<tr>
<td></td>
<td>217.06</td>
<td>0.83</td>
<td>0.05* 0.00** - 0.00** 0.00***</td>
</tr>
<tr>
<td>Duration data set</td>
<td>239.62</td>
<td>0.82</td>
<td>0.01* 0.07 0.00*** - 0.07</td>
</tr>
<tr>
<td></td>
<td>248.33</td>
<td>0.82</td>
<td>0.00** 0.16 0.00** 0.99 -</td>
</tr>
</tbody>
</table>

Note. *p < .05, **p < .01, ***p < .001.

Figure 2. Comparison of the pMSE and pMAPC results of different data sets

Figure 2 shows a box plot comparison of the different data sets based on the pMSE and pMAPC results. For each data set, we used box plots to describe the distribution of pMSE and pMAPC values for the 21 regression models obtained using PCR. The bottom and top lines represent the minimum and maximum values, respectively. From bottom to top, the three lines in the box indicate the lower quartile, median quartile, and upper quartile,
respectively. Figure 2 shows that the box plots of the duration data sets are longer than those of the accumulated data sets, which indicates that the predictive performance of the accumulated data sets was more stable than that of the duration data sets. In addition, the minimum pMSE values of the accumulated data sets are lower than those in the duration data sets and the maximum pMAPC values of the accumulated data sets are higher than those of the duration data sets. The results of the pMSE and pMAPC comparison show that the accumulated data sets have better prediction ability than do the duration data sets.

The results of the pMAPC and pMSE comparison matrix show that among the accumulated data sets, $W_{12}^{6}$ and $W_{10}^{6}$ had better predictive performance than did $W_{12}^{1}$. Compared with $W_{1}^{6}$, $W_{10}^{6}$ had a higher maximum value and higher median quartile for pMAPC, as well as a lower median quartile for pMSE. However, $W_{1}^{6}$ had the lowest pMSE value. These results show that $W_{12}^{10}$ had a slightly higher predictive performance and accuracy than did $W_{12}^{6}$. Because of outliers in the maximum value of pMSE and minimal value of pMAPC, the stability of $W_{1}^{10}$ was lower than that of $W_{1}^{6}$. In a real scenario, PCR would generate an equal number of regression results as variables of PCA. Thus, only one prediction result could be randomly selected from the results, which could cause issues if the data set had a wide range of prediction accuracy or in a data set with high average accuracy but few outliers such as $W_{12}^{6}$. Therefore, a convergent or stable data set is necessary even if its average accuracy is lower than that of other data sets. Thus, $W_{12}^{6}$ was determined to be the most suitable data set for real scenarios.

**Linear regression residual analysis**

According to the results of conventional regression and predictive performance estimation presented in the previous section, the accumulated data set $W_{12}^{6}$ had the highest stability and accuracy for predicting students’ final academic performance. A final test was required to identify the characteristics of normalization, independence, and homogeneity in the data set. However, because PCA can project data into a vector space with a dimension with the same number of variables, 21 models were estimated for each data set. To follow up $W_{12}^{6}$, we had to select the most predictable components from the 21 PCR results.

Figure 3 shows the pMSE and pMAPC results for each principle component in data set $W_{12}^{6}$. The optimal pMSE and pMAPC values (178.94 and 83.5%, respectively) can be obtained in the 12 components. Figure 4 shows the results of linear regression residual analysis by using a Q–Q plot of 12 principle components of $W_{12}^{6}$. The distribution for all residuals closely resembles a straight line, which indicates that the distribution for the difference between the predicted and real values supports the characteristics of normalization, independence, and homogeneity.

![Figure 3. Results of pMSE and pMAPC for each $W_{12}^{6}$ component](image)

To answer RQ1 (How early can we predict students’ final academic performance?), the results of the conventional and predictive performance estimations indicate that students’ final academic performance can be predicted by the sixth week of the semester. The PCR model from data set $W_{12}^{6}$ had the highest stability and prediction accuracy, which is consistent with the findings of previous studies, which achieved early identification of at-risk students after one third of the course period had been completed (Hu et al., 2014) and before the midterm exam (Romero et al., 2013). Data set $W_{12}^{10}$ had similar predictive accuracy and stability for predicting students’ final academic performance because performance can be calculated using quiz or homework scores throughout the whole semester. Hu et al. (2014) asserted that to identify at-risk students within the learning analytics framework, offering intervention based on an 18-week prediction result is too late. Therefore, the present study recommends using accumulated data set $W_{12}^{6}$ to predict students’ final academic performance. In
addition, we found that the predictive performance of duration data sets is inferior to that of accumulated data sets, which indicates that the completeness of data collection is crucial for data analysis.

**Determining critical factors that affect students' final academic performance in blended learning**

According to the summary of the literature review, the first step to predicting students’ final academic performance is to determine as many variables as possible. Subsequently, rules should be applied to enable the selection of variables to obtain higher prediction ability. Moreover, according to the summary in previous section, data set $W_1$ had the highest stability and predictive accuracy, and thus we used this data set to determine the critical factors that affect students’ learning performance. Table 8 shows the regression model estimation results. Components 1, 2, 5, 7, 9, 10, and 12 had a significant influence on students’ final academic performance. For each significant component, we selected variables with higher coefficients as critical factors; for example, variable $X_2$ was selected as the critical factor for Component 1 because of the substantial differences between the coefficient of variable $X_2$ and those of the other variables.

**Table 8. Variable estimation results of PCR for 12 components obtained using data set $W_1$**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Components 1</th>
<th>Components 2</th>
<th>Components 3</th>
<th>Components 4</th>
<th>Components 5</th>
<th>Components 6</th>
<th>Components 7</th>
<th>Components 8</th>
<th>Components 9</th>
<th>Components 10</th>
<th>Components 11</th>
<th>Components 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>0</td>
<td>-0.01</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>-0.01</td>
<td>0.01</td>
<td>0.04</td>
<td>0.18</td>
<td>-0.14</td>
<td>0.21</td>
<td>0.01</td>
</tr>
<tr>
<td>$X_2$</td>
<td><strong>0.99</strong></td>
<td>-0.17</td>
<td>0.01</td>
<td>0.03</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$X_3$</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
<td>-0.01</td>
<td>0.02</td>
<td>0.03</td>
<td>-0.02</td>
<td>-0.06</td>
<td>0.16</td>
<td>-0.03</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>$X_4$</td>
<td>0.01</td>
<td>0.03</td>
<td>-0.13</td>
<td>0.2</td>
<td>0.34</td>
<td>-0.2</td>
<td>-0.08</td>
<td>0.5</td>
<td>0.12</td>
<td>0.03</td>
<td>0.14</td>
<td>0.17</td>
</tr>
<tr>
<td>$X_5$</td>
<td>0.01</td>
<td>0.05</td>
<td>-0.12</td>
<td>0.23</td>
<td>0.35</td>
<td>-0.13</td>
<td>0.08</td>
<td>-0.19</td>
<td>0.04</td>
<td>0.09</td>
<td>0.08</td>
<td>0.10</td>
</tr>
<tr>
<td>$X_6$</td>
<td>0.01</td>
<td>0</td>
<td>-0.02</td>
<td>-0.08</td>
<td>0.15</td>
<td>0</td>
<td>-0.14</td>
<td>-0.15</td>
<td>0.37</td>
<td>-0.28</td>
<td>-0.12</td>
<td>-0.36</td>
</tr>
<tr>
<td>$X_7$</td>
<td>0.01</td>
<td>0.03</td>
<td>-0.14</td>
<td>0.05</td>
<td>0.14</td>
<td>-0.1</td>
<td>-0.05</td>
<td>0.38</td>
<td>0.03</td>
<td>-0.05</td>
<td>0.33</td>
<td>0.10</td>
</tr>
<tr>
<td>$X_8$</td>
<td>0</td>
<td>-0.01</td>
<td>0.05</td>
<td>-0.12</td>
<td>0.24</td>
<td>0.36</td>
<td>-0.14</td>
<td>0.09</td>
<td>-0.26</td>
<td>-0.04</td>
<td>0.04</td>
<td>-0.02</td>
</tr>
<tr>
<td>$X_9$</td>
<td>0.16</td>
<td><strong>0.95</strong></td>
<td>0.26</td>
<td>0.06</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.03</td>
<td>0.02</td>
<td>0</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$X_{10}$</td>
<td>0.06</td>
<td>0.26</td>
<td>-0.94</td>
<td>-0.17</td>
<td>-0.06</td>
<td>0.12</td>
<td>0.03</td>
<td>0</td>
<td>-0.01</td>
<td>-0.03</td>
<td>0.01</td>
<td>0</td>
</tr>
<tr>
<td>$X_{11}$</td>
<td>0.02</td>
<td>0.03</td>
<td>-0.08</td>
<td>-0.24</td>
<td><strong>0.63</strong></td>
<td>-0.65</td>
<td>-0.29</td>
<td>-0.04</td>
<td>-0.09</td>
<td>-0.04</td>
<td>0.09</td>
<td>0.07</td>
</tr>
<tr>
<td>$X_{12}$</td>
<td>0.03</td>
<td>0.01</td>
<td>0.19</td>
<td>-0.9</td>
<td>-0.29</td>
<td>-0.06</td>
<td>0.22</td>
<td>-0.05</td>
<td>-0.04</td>
<td>-0.05</td>
<td>0.01</td>
<td>-0.05</td>
</tr>
<tr>
<td>$X_{13}$</td>
<td>0.01</td>
<td>-0.01</td>
<td>0.06</td>
<td>-0.12</td>
<td>0.3</td>
<td>0.39</td>
<td>-0.17</td>
<td>0.15</td>
<td>-0.29</td>
<td>-0.06</td>
<td>-0.06</td>
<td>-0.22</td>
</tr>
<tr>
<td>$X_{14}$</td>
<td>0</td>
<td>-0.02</td>
<td>0</td>
<td>-0.02</td>
<td>-0.21</td>
<td>-0.09</td>
<td>-0.31</td>
<td>0.67</td>
<td>0.24</td>
<td>-0.42</td>
<td>0.25</td>
<td>0</td>
</tr>
<tr>
<td>$X_{15}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-0.01</td>
<td>-0.07</td>
<td>-0.03</td>
<td>-0.13</td>
<td>0.12</td>
<td>0</td>
<td>0.06</td>
<td>-0.06</td>
<td>0.05</td>
</tr>
<tr>
<td>$X_{16}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-0.05</td>
<td>-0.01</td>
<td>-0.07</td>
<td>0.08</td>
<td>-0.02</td>
<td>0.02</td>
<td>-0.08</td>
<td>-0.02</td>
</tr>
<tr>
<td>$X_{17}$</td>
<td>0</td>
<td>-0.01</td>
<td>0</td>
<td>-0.02</td>
<td>-0.1</td>
<td>-0.03</td>
<td>-0.14</td>
<td>0.15</td>
<td>0.02</td>
<td>0.04</td>
<td>-0.11</td>
<td>0</td>
</tr>
<tr>
<td>$X_{18}$</td>
<td>0</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.07</td>
<td>-0.31</td>
<td>-0.09</td>
<td><strong>0.49</strong></td>
<td>0.2</td>
<td>-0.21</td>
<td><strong>0.62</strong></td>
<td>-0.22</td>
<td>0.08</td>
</tr>
<tr>
<td>$X_{19}$</td>
<td>0</td>
<td>0.01</td>
<td>0</td>
<td>-0.17</td>
<td>0</td>
<td><strong>0.41</strong></td>
<td>-0.36</td>
<td>-0.03</td>
<td><strong>0.55</strong></td>
<td>-0.59</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$X_{20}$</td>
<td>0</td>
<td>-0.01</td>
<td>0.01</td>
<td>0.03</td>
<td>-0.27</td>
<td>0.04</td>
<td><strong>0.43</strong></td>
<td>-0.5</td>
<td>-0.13</td>
<td>-0.01</td>
<td>0.65</td>
<td>-0.11</td>
</tr>
<tr>
<td>$X_{21}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-0.02</td>
<td>0.03</td>
<td>0.01</td>
<td>-0.09</td>
<td>-0.1</td>
<td>0.01</td>
<td>0.04</td>
<td><strong>0.81</strong></td>
</tr>
</tbody>
</table>

$p$ value $0^{***}$ $0.009^{**}$ 0.881 0.637 $0.02^*$ 0.81 $0.006^{**}$ 0.114 $0.033^*$ $0.001^{**}$ 0.099 $0.003^{**}$

Note: $^{*} p < .05$, $^{**} p < .01$, $^{***} p < .001$.

To address RQ2 (Which are the most critical factors that affect students’ final academic performance in blended learning?), this study determined seven critical factors that affect students’ final academic performance, namely...
X₂ (Number of activities a student engages in per week), X₆ (Number of times a student clicks “Play” during video viewing per week), X₁₁ (Number of times a student clicks “Backward seek” during video viewing per week), X₁₈ (Student’s weekly practice score), X₁₉ (Student’s weekly homework score), X₂₀ (Student’s weekly quiz score), and X₂₁ (Number of times a student participates in after-school tutoring per week).

X₁₈, X₁₉, and X₂₀ are critical factors that affect students’ final academic performance because of the evident relationships between each of these three variables and learning performance. The results are consistent with the findings of Huang and Fang (2013), who determined that exam scores and homework scores can predict students’ final academic performance. Xing et al. (2016) asserted that online learning behaviors can predict dropout only in online courses. Based on our identification of four online variables, X₂, X₅, X₁₁, and X₁₈, as critical factors that affect students’ final academic performance, dropout and students’ final academic performance may be related.

Ability of different data sets (blended vs. online vs. traditional) to predict students’ final academic performance in blended learning

As mentioned in previous section, we identified seven critical factors that affect students’ final academic performance in MOOC and OAS enabled blended courses. These seven critical factors can be categorized in W₇ as blended, online, and traditional data sets. These seven critical factors can be categorized in W₇ (critical factors) as blended, online, and traditional data sets. Table 9 lists the categories of each factor and the PCR results. W₇(O), W₇(T), and W₇(B) represent online, traditional, and blended data sets, respectively.

The results of R², the F-test, and the Durbin–Watson test, demonstrate that each indicator was acceptable for each data set (Table 9). The independent variables in three data sets are listed in Table 9. The regression tests for W₇(O), W₇(T), and W₇(B) contained three, three, and five significant variables, respectively, which indicates that the selected critical factors are crucial for predicting students’ final academic performance. In addition, the numbers of best components for the online, traditional, and blended data sets were all equal to the numbers of independent variables for each data set, which shows that each data set required whole independent variables to determine the optimal predictive performance. The blended data set W₇(B) obtained the optimal pMSE and pMAPC values of 159.17 and 0.82, respectively. Figure 3 illustrates that the optimal pMSE in W₇(B) was 178.94, which was inferior to that of blended dataset W₇(B). These results show that the selected critical factors not only reduce the number of variables for PCR but also improve prediction performance.

<table>
<thead>
<tr>
<th>Data set</th>
<th>Variables (critical factors)</th>
<th>p value</th>
<th>pMSE</th>
<th>pMAPC</th>
<th>Best Comp</th>
<th>R²</th>
<th>F</th>
<th>Durbin–Watson</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data set which blended online and traditional critical factors W₇(B)</td>
<td>X₂</td>
<td>0.00&lt;sup&gt;**&lt;/sup&gt;</td>
<td>159.17</td>
<td>0.82</td>
<td>7 (DF = 7)</td>
<td>0.56</td>
<td>0.00&lt;sup&gt;***&lt;/sup&gt;</td>
<td>1.62</td>
</tr>
<tr>
<td></td>
<td>X₉</td>
<td>0.01&lt;sup&gt;**&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>X₁₁</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>X₁₈</td>
<td>0.00&lt;sup&gt;**&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>X₁₉</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>X₂₀</td>
<td>0.11&lt;sup&gt;**&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>X₂₁</td>
<td>0.01&lt;sup&gt;*&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data set of online critical factors W₇(O)</td>
<td>X₂</td>
<td>0.00&lt;sup&gt;**&lt;/sup&gt;</td>
<td>181.16</td>
<td>0.82</td>
<td>4 (DF = 4)</td>
<td>0.39</td>
<td>0.00&lt;sup&gt;***&lt;/sup&gt;</td>
<td>1.42</td>
</tr>
<tr>
<td></td>
<td>X₉</td>
<td>0.03&lt;sup&gt;*&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>X₁₁</td>
<td>0.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>X₁₈</td>
<td>0.00&lt;sup&gt;**&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data set of traditional critical factors W₇(T)</td>
<td>X₁₉</td>
<td>0.00&lt;sup&gt;**&lt;/sup&gt;</td>
<td>186.99</td>
<td>0.80</td>
<td>3 (DF = 3)</td>
<td>0.40</td>
<td>0.00&lt;sup&gt;***&lt;/sup&gt;</td>
<td>1.70</td>
</tr>
<tr>
<td></td>
<td>X₂₀</td>
<td>0.00&lt;sup&gt;**&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>X₂₁</td>
<td>0.03&lt;sup&gt;*&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .05, **p < .01, ***p < .001.

To answer RQ3 (Which type of data set (blended vs. online vs. traditional) is more effective for predicting students’ final academic performance in blended learning?), the blended data set obtained the most favorable predictive performance, demonstrating that the blended data set had a higher predictive performance than did the traditional data set. This result is consistent with the findings of Agudo-Peregrina et al. (2014), who revealed that students’ interactions with online learning environments influence their academic performance. In addition, the present study followed previous studies in using critical factors to improve predictive performance (Asif et al., 2014; Romero et al., 2013; Villagrá-Arnedo et al., 2016).
Conclusion

This study collected student profiles from a MOOC and OAS enabled blended Calculus course. In addition, we applied PCR to evaluate five data sets that were separated based on the collected data. The experimental results demonstrate that students’ final academic performance in a blended Calculus course can be predicted with high stability and accuracy by a data set containing data from the first two weeks. In other words, through well-identified online and traditional variables, we were able to predict students’ final academic performance when the balance reached one-third of the way through the semester. Seven critical factors that influence students’ learning performance were identified by the regression model to improve prediction performance. However, explaining the relationship between these critical factors and learning performance would require investigation through interviews with educational experts. Furthermore, to achieve the goal of improving students’ learning performance, the student performance prediction model proposed in this study and a well-defined intervention strategy must be integrated into the learning analytics framework. The complete learning analytics framework could be applied to predict student learning outcomes in the second semester of such a Calculus course.

Acknowledgments

This work was supported by Ministry of Science and Technology, Taiwan under grants MOST-104-2511-S-008-006-MY2, MOST-105-2511-S-008-003-MY3, MOST-105-2622-S-008 -002-CC2, MOST-106-2511-S-008 -004 -MY3.

References


Personalized Word-Learning based on Technique Feature Analysis and Learning Analytics

Di Zou¹ and Haoran Xie²

¹Department of English Language Education, The Education University of Hong Kong, Hong Kong // ²Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong // dizoudaisy@gmail.com // hrxie2@gmail.com

Corresponding author

ABSTRACT
Many studies have highlighted the importance of personalized learning, and most current e-learning systems are able to personalize materials, activities, etc., based on individualized learner-factors. However, none of the extant word-learning systems provides a personalized learning experience that is guided by a comprehensive word learning theory. In this study, we develop such a system based on Nation and Webb’s checklist for technique feature analysis - a thorough set of factors that promote effective word learning. This system recommends personalized word learning tasks based on the technique feature analysis scores of different tasks and user models. To examine the effectiveness of the proposed system, we conducted an experiment among 105 English learners, grouped them into three teams randomly, and asked them to learn forty target words through three approaches: a non-personalized approach, a personalized approach guided by a partial version of the technique feature analysis, and a personalized approach guided by the full list of the technique feature analysis. Significant differences were observed among the effectiveness of the three approaches in promoting word learning, with the personalized approach guided by the complete checklist leading to the best learning performance. It is therefore suggested that e-learning systems should be designed based on comprehensive learning theories.

Keywords
Personalized learning, Vocabulary acquisition, Learning analytics, Technique feature analysis, User model

Introduction and background

With the rapid development of information technology and pervasive use of digital devices in recent years, language education is steadily moving toward utilizing more technology-based tools and approaches. Computer-assisted and Mobile-Assisted Language Learning (CALL, MALL) have gained increasing popularity, and a large number of language-enhancement systems have emerged. For example, Chen and Hsu (2008) developed a personalized intelligent mobile-learning system which recommends appropriate English news articles to learners based on their reading abilities as evaluated by a fuzzy item response theory proposed in the research. Liu (2009) built a ubiquitous learning system - the Handheld English Language Learning Organization (HELLO) - which integrates augmented reality and sensors for supporting learners’ development of listening and speaking skills. Similarly, Kwon et al. (2010) presented a personalized computer-assisted language learning system based on learners’ cognitive abilities with respect to their language proficiency levels; this system employed a strategy of retrieval learning, a method of learning memory cycle, and a method of repeated learning for improving learning effectiveness. Focusing on learning environments, Wu, Sung, Huang, Yang and Yang (2011) developed a situated and reading-based English learning system that integrated a reading guidance mechanism into the development of an e-learning environment. Moreover, Hsieh, Wang, Su and Lee (2012) designed a fuzzy logic-based personalized learning system to support adaptive English learning. Hsu, Hwang and Chang (2013) developed a personalized recommendation-based mobile learning approach to improving reading skills of language learners. Additionally, Xie et al. (2017) proposed a profile-based approach to discovering learning paths for group users to improve the learning effectiveness and efficiency of a whole group. A common feature of all aforementioned systems and approaches is that they provide a personalized learning experience based on the needs, prior knowledge, preferences and/or learning styles of individual learners, and thus are more effective than non-personalized systems.

With close connections to grammatical knowledge and other language skills, word knowledge is widely acknowledged by linguists and teachers as the foundation of language acquisition (Lightbown & Spada, 2006). Language learners also attribute great importance to word knowledge and are particularly interested in effective approaches to word learning, because they believe that word knowledge is central to their communicative competence (Schmitt, 2000). However, many learners feel frustrated about vocabulary learning because they often forget words that they have previously learned. They also do not know what words should be learned first and how to learn them effectively. The probability of knowing words through implicit learning is very small, whereas it is difficult to engage in explicit learning for a long period of time (Nation, 2001), a large number of
learners therefore regard word learning as a time-consuming, yet not necessarily rewarding, activity. Identifying effective word learning methods or techniques is therefore paramount.

**Personalized learning systems for vocabulary learning**

Given the importance of word knowledge for language acquisition and students’ preference for greater use of digital technological tools in education (Zou & Lambert, 2017), numerous e-learning systems are specifically designed for supporting effective vocabulary learning. Most of them provide a personalized learning experience for the users, as learning vocabulary by using desired and relevant language resources is essential for effective acquisition (Zou et al., 2017). A few representative studies are reviewed as follows.

Barker (2007) proposed a personalized approach that enables language learners to make their own decisions about the costs and benefits of learning new words through analyzing these words via consideration of both word- and learner-specific factors when they encounter the words. Jung and Graf (2008) developed a word association game to facilitate personalized vocabulary learning in a web-based system; this system can effectively increase the motivation of learners and cater to their individual needs. Moreover, Chen and Chung’s (2008) personalized mobile English vocabulary learning system, which is based on item response theory and learning memory cycle, can appropriately recommend vocabulary for learning according to individual learners’ word knowledge and memory cycle. Chen and Li (2010) also designed a personalized context-aware ubiquitous vocabulary learning system based on learner locations as detected by wireless positioning techniques, learning time, individual English vocabulary abilities, and leisure time. This system enables learners to adapt their learning content to effectively support English vocabulary learning in a school environment. Furthermore, Huang, Huang, Huang and Lin (2012) developed an easy-to-use ubiquitous English vocabulary learning system to assist students in experiencing a systematic vocabulary learning process. Employing fuzzy inference mechanisms, memory cycle updates, learner preferences and analytic hierarchy processes, Hsieh et al. (2012) proposed a personalized English article recommendation system which selects appropriate articles for learners by using accumulated learner profiles. This system can effectively improve learners’ English proficiency levels in an extensive reading environment via helping them comprehend new words quickly and review words that they knew implicitly. Similarly, the mobile learning system of Hsu et al. (2013) includes a reading material recommendation mechanism that suggests articles to learners based on their preferences and knowledge levels. It also involves a reading annotation module that enables students to take notes of English vocabulary translations for the reading content in individual or shared annotation mode. Sandberg, Maris and Hoogendoorn (2014) compared the learning performance of two groups of learners who participated in a mobile learning application, and noted that gaming contexts and intelligent adaptation have additional value for mobile vocabulary learning.

Zou, Xie, Li, Wang and Chen (2014) presented a personalized word learning task recommendation system based on Laufer and Hulstijn’s (2001) involvement load hypothesis, and found that this system which personalizes learners’ learning experience via load-based learner profiles promotes effectively word learning. Additionally, Wang and Shih (2015) examined the effects of self-paced use of smart phones as mobile learning tools on English vocabulary learning, and found that the group with mobile learning scored significantly higher than the control group. Likewise, Huang, Yang, Chiang and Su (2016) investigated the effects of a situated mobile learning approach on students’ English learning motivation and performance by developing a five-step vocabulary learning strategy and a mobile learning tool in a situational vocabulary learning environment. The research findings showed that the proposed strategy and tool are effective in increasing students’ learning motivation and performance. With similar purposes, but a different theoretical framework, Xie, Zou, Lau, Wang and Wong (2016) developed an e-learning system for recommending vocabulary learning tasks based on topic-based profiles obtained from social media platforms and load-based profiles measured by the involvement load hypothesis. The experiment results demonstrated that the proposed system not only improves learning effectiveness, but also increases learning enjoyment.

However, most of the aforementioned personalized vocabulary learning systems involve just one or two factors that are conducive to effective word learning, and thus are limited in other respects. For example, Chen and Chung’s (2008) system recommends appropriate English vocabulary for learning according to different learners’ prior word knowledge and memory cycle, and thus it is effective in adjusting learning modes of various learners to promote their learning performances and interests. Nevertheless, this system promotes little development of productive word knowledge, as no generative use is involved in the learning experience. Similarly, although the system designed by Huang et al. (2016) takes into account facilitative factors for word learning, such as high motivation, retrieval of words and meaningful contexts, it is limited in terms of the spacing between retrievals, linking of form and meaning, and generative use of the words. This is probably because the design of these systems is not guided by a comprehensive checklist that covers all important factors that are essential for
effective vocabulary learning. Therefore, it is necessary to develop a vocabulary learning system under the umbrella of a comprehensive set of word learning techniques. Nation and Webb’s (2011) checklist for technique feature analysis (hereafter, TFA) is selected as the theoretical framework because it constitutes an elaborate set of criteria which provides a reliable guide for predicting, evaluating, and explaining the effectiveness of diverse word-focused tasks.

The checklist for technique feature analysis

Nation and Webb’s (2011) checklist for technique feature analysis, which operationalizes cognitive notions, such as depth of processing and richness of encoding, involves five main components: motivation, noticing, retrieval, generation, and retention. Each of these five main components further includes three to five questions, covering various factors that are effective in promoting word learning. There are altogether 18 questions in the checklist, and point values are used to evaluate different word learning techniques (see Table 1).

<table>
<thead>
<tr>
<th>Component</th>
<th>Criteria</th>
<th>Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivation</td>
<td>Is there a clear vocabulary learning goal?</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Does the activity motivate learning?</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Do the learners select the words?</td>
<td>0</td>
</tr>
<tr>
<td>Noticing</td>
<td>Does the activity focus attention on the target words?</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Does the activity raise awareness of new vocabulary learning?</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Does the activity involve negotiation?</td>
<td>0</td>
</tr>
<tr>
<td>Retrieval</td>
<td>Does the activity involve retrieval of the word?</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Is it productive retrieval?</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Is it recall?</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Are there multiple retrievals of each word?</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Is there spacing between retrievals?</td>
<td>0</td>
</tr>
<tr>
<td>Generation</td>
<td>Does the activity involve generative use?</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Is it productive?</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Is there a marked change that involves the use of other words?</td>
<td>0</td>
</tr>
<tr>
<td>Retention</td>
<td>Does the activity ensure successful linking of form and meaning?</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Does the activity involve instantiation?</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Does the activity involve imaging?</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Does the activity avoid interference?</td>
<td>0</td>
</tr>
<tr>
<td>Maximum score</td>
<td></td>
<td>18</td>
</tr>
</tbody>
</table>

Specifically, the component “motivation” concerns whether an activity has a clear word learning goal, whether it motivates learners, and whether the learners select the words. The component “noticing” questions whether an activity induces attention on the target words, whether awareness of new word learning is raised, and whether negotiation is involved. The component “retrieval” consists of receptive retrieval, productive retrieval, recall, multiple retrievals, and spacing between retrievals. The component “generation” comprises generative use, productive generation, and marked changes that involve the use of other words. The component “retention” mainly refers to whether an activity ensures linking of form and meaning, and whether it involves instantiation, imaging, and avoids interference (Nation & Webb, 2011). Scores of the questions are measured in a binary manner: one will be given if a learning activity involves a certain criterion, and zero will be given if not.

Taking the task reading comprehension and performing cloze-exercises with textual annotations as an example, its total TFA score is 7. As demonstrated in Table 2, because this task requires the learners to compare different words and fill them in the blanks where the contexts are suitable for the words, it has a clear word learning goal with focused attention on the target words. Moreover, as the students are aware of the learning of the target words while matching them with the appropriate contexts, receptive generative use of these words is induced, and the students are motivated. The generative use of the target words induced here is not productive since the contexts are given, and the students do not generate original contexts themselves. However, as the students need to fill in the blanks with the target words, the activity ensures successful linking of form and meaning. Lastly, since the words are normally not members of the same lexical set, the activity avoids interference. The total TFA score of cloze-exercises is therefore 7. Compared to cloze-exercises, writing original sentences using target words has two more TFA scores because learner-created original contexts for the target words are generated, and hence productive generative use of target words and marked changes that involve the use of other words are induced. Except for this point, these two tasks are similar in other aspects of the TFA criteria. They both induce
clear learning goals, raise awareness of word learning, draw attention to the target words, ensure linking of form and meaning, motivate learning, and avoid interference (see Table 2).

<table>
<thead>
<tr>
<th>Component</th>
<th>Criteria</th>
<th>Scores of two tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Cloze-exercises</td>
</tr>
<tr>
<td>Motivation</td>
<td>Is there a clear vocabulary learning goal?</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Does the activity motivate learning?</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Do the learners select the words?</td>
<td>0</td>
</tr>
<tr>
<td>Noticing</td>
<td>Does the activity focus attention on the target words?</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Does the activity raise awareness of new vocabulary learning?</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Does the activity involve negotiation?</td>
<td>0</td>
</tr>
<tr>
<td>Retrieval</td>
<td>Does the activity involve retrieval of the word?</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Is it productive retrieval?</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Is it recall?</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Are there multiple retrievals of each word?</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Is there spacing between retrievals?</td>
<td>0</td>
</tr>
<tr>
<td>Generation</td>
<td>Does the activity involve generative use?</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Is it productive?</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Is there a marked change that involves the use of other words?</td>
<td>0</td>
</tr>
<tr>
<td>Retention</td>
<td>Does the activity ensure successful linking of form and meaning?</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Does the activity involve instantiation?</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Does the activity involve imaging?</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Does the activity avoid interference?</td>
<td>1</td>
</tr>
<tr>
<td>Total score</td>
<td></td>
<td><strong>7</strong></td>
</tr>
</tbody>
</table>

According to Nation and Webb (2011), tasks with higher TFA scores promote better word learning than tasks with lower scores. Hu and Nassaji’s (2016) study also provided empirical evidence for the reliability of TFA in evaluating the effectiveness of diverse word learning tasks. This article also noted that TFA has good explanatory power in predicting word learning gains. In the present study, we develop a personalized vocabulary learning system under the umbrella of the checklist. The detailed contributions of this article are listed as follows.

- A user model is developed with the checklist for technique feature analysis as its theoretical framework, and thus a comprehensive set of factors that facilitates effective word learning is covered.
- Personalized vocabulary learning processes are generated by the proposed system based on the user model to assist learners to select appropriate learning tasks.
- Real participants are invited to use the proposed system, and the effectiveness of this personalized task recommendation system is verified by the learning performance of the participants.
- Implications of the research findings are elaborated from the perspectives of how technology-enhanced word learning tasks and personalized e-learning systems should be designed.

The remaining sections of this article are organized as follows. The methodology section describes the development of the user model and the generation of personalized vocabulary learning. The experiment and results sections explain the settings, processes, and results of the experiment. The implications and conclusion of the study and future research directions are discussed in the last section.

**Methodology**

In this section, we will introduce the method of the study, focusing on the development of the user model and the generation of personalized vocabulary learning.

**User modelling based on the checklist for technique feature analysis**

As mentioned previously, the theoretical foundation of the user model in the proposed vocabulary learning system is the checklist for technique feature analysis. To measure the TFA score of a word, the TFA scores of all tasks that include this word as a target word will be counted. For example, if learner A has completed two tasks
(i.e., task A and task B) that focus attention on the word B, the TFA score of word B for learner A is the sum of the TFA scores of task A and task B, as shown in Figure 1. Formally, a user model is defined as a matrix of TFA scores of each word with respect to the 18 TFA criteria as follows:

\[ L_i = (s_{mn}) \in \mathbb{R}^{m \times n} \]  

(1)

where \( L_i \) is the user model for learner \( i \); \( s_{mn} \) is an entry in \( m \)-th row and \( n \)-th column; \( m \) is the size of the collection of target words in the system; and \( n \) is the total number of criteria in the TFA (i.e., \( n = 18 \)).

To calculate the TFA score of each entry, the scores of all tasks that induce learning of the target word are considered. Specifically, the overall TFA score of entry \( s'_{xy} \) is the total TFA scores of all tasks that include the target word \( x \) for criteria \( y \) as follows:

\[ s'_{xy} = \sum_{t \in T_x^i} s_y(t) \]  

(2)

where \( T_x^i \) is the set of all tasks, which include word \( x \) as a target word, studied by learner \( i \); and \( s_y(t) \) denotes the binary scoring function, as shown in Table 1. A larger value of \( s'_{xy} \) implies that learner \( i \) has a higher load of learning on the target word \( x \) in terms of criteria \( y \).

The generation of personalized vocabulary learning

Personalization is a learning technique which has been widely adopted by various disciplines, including natural science (Hwang, Kuo, Yin, & Chuang, 2010; Hwang, Sung, Hung, Huang, & Tsai, 2012), mathematics (Chen & Liu, 2007), and management (Xie et al., 2017). The user model based on the TFA, as explained above, enables the system to track the TFA scores of all target words in each criterion during a user’s learning process. To better exploit such a user model, it is important to track the learning processes of every learner and then generate personalized learning to improve his or her learning effectiveness. The generation of a personalized learning process in this learning system is achieved via two approaches: TFA utility and task diversity.
TFA utility

The TFA checklist is utilized from two aspects by the learning system when recommending a task to a learner. Firstly, it is used to evaluate the effectiveness of several candidate tasks, and one is selected that can best promote the learning of target words. The selection of the task follows the principle that the new task induces components of the TFA criteria that are different from, but related to, the previous tasks. This aims to help users learn various aspects of knowledge of the target words via performing different tasks, and practice learning them in different ways. For example, if Task 1 focuses on the component “retention” and involves imaging and linking of form and meaning, Task 2 may then emphasize the component “generation” by asking learners to create original contexts for the target words. Moreover, Task 3 may impose retrieval of the previously learned target words as guided by the component “retrieval.”

Therefore, the first part of the TFA utility is formally defined as follows:

\[
util_1(t) = \sum_{s \in C} s_y(t)
\]

where \(C\) is the set of all TFA criteria. Based on these principles, the system recommends a task with maximal utility for learner \(i\).

The second aspect of the TFA utility enables the system to recommend a task that can assist a learner to increase the TFA scores of entries with lower scores in the user model. This is achieved by recommending tasks that emphasize the TFA components with lower scores as induced by previous learning tasks. In this way, the TFA scores in the user model can be “fully utilized” after the participants complete the current learning task. Formally, the second part of the TFA utility is defined as follows:

\[
util_2(t, L_i) = \sum_{y=1}^{n} \sum_{x=1}^{m} \frac{\Delta s_{xy}^i}{1 + s_{xy}^i}
\]

where \(s_{xy}^i\) is an entry of the TFA score in the user model; \(\Delta s_{xy}^i\) is the change of the TFA score of this entry after the completion of task \(t\); and \(1 + s_{xy}^i\) is to avoid a zero score for the denominator. The core principle here is to find a learning task that can increase the scores of more entries with low TFA scores in the user model. The priority is to fill entries with low TFA scores. The overall utility score is consolidated by using the following aggregation method:

\[
util(t, L_i) = e^{util_1(t)} \times e^{util_2(t, L_i)}
\]

where \(util_1()\) and \(util_2()\) are the two functions as defined in the equations (3) and (4), respectively. The proposed learning system basically prioritizes tasks with larger overall utility and recommends them to users.

Task diversity

Task diversity refers to the degree to which a task is different from previous tasks. For two tasks \(A\) and \(B\) which have a similar TFA utility, task \(A\) can better motivate a learner if it is more dissimilar from previous learning tasks than \(B\). The principle here, which is suggested by Nation (2001) and partially supported by a study (Xie et al., 2016), is that tasks with greater diversity lead to better word retention than tasks of similar types. Formally, the diversity of a task \(t\) to the set of previously learned tasks \(T^i\) is defined as follows:

\[
div(t, T^i) = 1 - \frac{1}{|T^i|} \sum_{t \in T^i} \frac{v^t \cdot v^T}{\|v^t\| \cdot \|v^T\|}
\]

where \(v^t\) is a vector of TFA scores with five components (i.e., \(v^t \in \mathbb{R}^5\)) for the task \(t\); \(v^T\) is a previous task in the set \(T^i\); and \(|T^i|\) is the total number of previous learning tasks. The overall diversity function \(div()\) measures the dissimilarity of a task \(t\) to the set of previous learning tasks \(T^i\) by calculating the inverse of the average cosine similarity.
Table 3. An example of calculating task diversity in different granularities

<table>
<thead>
<tr>
<th>Component</th>
<th>Criteria</th>
<th>(t_a)</th>
<th>(t_b)</th>
<th>(t_c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noticing</td>
<td>Does the activity involve generative use?</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Is it productive?</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Is there a marked change that involves the use of other words?</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Retention</td>
<td>Does the activity ensure successful linking of form and meaning?</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Does the activity involve instantiation?</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Does the activity involve imaging?</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Does the activity avoid interference?</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Note that we adopt a vector of the five main components of the TFA rather than all 18 criteria when calculating the degree of diversity, because the main components have better granularity than the sub-criteria in terms of deciding how similar two learning tasks are. Suppose that there are three tasks \(t_a\), \(t_b\) and \(t_c\), as shown in Table 3. For simplicity, only seven criteria from two main components are displayed in this example. If the criteria are used as the vector dimension for measuring diversity, the diversity between \(t_a\) and \(t_b\) is the same as that between \(t_a\) and \(t_c\). However, this is unreasonable, as \(t_c\) compared to \(t_b\) is more different from \(t_a\). The TFA scores of \(t_a\) and \(t_b\) are mainly in the component “retention”, while the TFA scores of \(t_a\) are mainly in the category “noticing”. Thus, the degree of diversity is more reasonable when measured from the main component level (i.e., \(div(t_a, t_c)\) is larger than \(div(t_a, t_b)\)). Therefore, in our model, the diversity among different tasks is measured from the perspective of the main components, rather than the sub-criteria of the TFA.

Moreover, as each component of the TFA has different scales of scores (e.g., scores of “noticing” range from 0 to 3 and scores of “retention” range from 0 to 4), the score of each main component of a task \(c(t)\) is normalized in the model by using the following method:

\[
c_j(t) = \frac{\sum_{s_j \in c_j} s_j(t)}{|c_j|}
\]

where \(c_j\) is a component which includes several criteria; \(\sum_{s_j \in c_j} s_j(t)\) is to aggregate the TFA scores under a component; and \(|c_j|\) is the number of criteria under a component.

Learning process generation

The learning process is generated in an interactive manner. As shown in Figure 2, the detailed steps of the whole interactive process are as follows.

![Figure 2. The interactive process of the system](image)
Step 1: The learner selects two tasks as the starting point;
Step 2: The system suggests two tasks to the learner based on the initial task set in terms of TFA utility and task diversity;
Step 3: The learner selects one of the two recommended tasks and continues learning;
Step 4: The system updates the learning history of the learner and goes back to step 2.

An illustrative example of the overall generic framework

To better illustrate the overall generic framework, the learning process of an example user is shown in Figure 3. The learning logs record all relevant information about the tasks that the user has completed, including the task id, time, task types, target words, and the TFA scores of the tasks. Taking Task 7 as an example, it was completed by the learner on April 23, 2017 at 16:15; it is cloze-exercises, and the target words include renge, trait, etc. The total TFA score of this task is 7, and the sub-scores of the 18 TFA criteria are individual (1, 1, 0, 1…) as listed in brackets. The user model is a matrix as defined in Equation (1). Each entry of the matrix denotes the history record of a participant’s learning of a target word in terms of one of the 18 TFA criteria. For example, entry 4 of the first column of the first row denotes that the first target word has been learned four times by this user through performing four tasks with clear vocabulary learning goals. According to the learning logs and user models, the system suggests learning tasks based on the TFA utility and task diversity as respectively defined in Equation (5) and Equation (6). Normally, the suggested learning tasks: (1) are of different types as compared to the tasks that have been completed by the user recently; and (2) tend to focus more on the target words that the user has not encountered through performing other tasks. In each cycle, when the user has selected and completed a task, the system updates the user’s learning logs and user models automatically, and based on the latest data, the system further makes recommendations of new learning tasks. As the system continues to iterate, the user’s learning experience throughout the entire learning process is personalized with respect to his or her learning logs and user models, and the recommended tasks are guided by TFA utility and task diversity.

![Figure 3. An illustrated example of the overall generic framework](image)

Experiments

Subjects

To verify the effectiveness of the proposed system, real subjects were invited to participate in the experiments. A total of 105 students from universities in Hong Kong and mainland China participated in the study. Their ages ranged from 18 to 28, and a wide variety of majors were covered, including business studies, engineering, humanities, social sciences, biology, medicine, and physical sciences (see Table 4).

These participants were randomly assigned into three groups. Each group adopted a different learning approach to using the word-learning task recommendation system. Group A adopted a non-personalized learning approach, while Group B and Group C adopted personalized learning approaches. The main difference between the approaches adopted by Group B and Group C is that the generation of personalized learning processes for the participants of Group B was guided by an earlier version of the TFA, where only three main components: noticing, retrieval, and generation are included; whereas, the generation of personalized learning processes for the participants of Group C was guided by the complete checklist of TFA, as shown in Table 1. The earlier version of the TFA does not quantify features concerning elaboration, while the current checklist of TFA covers
such features and increases the number of elaboration parameters (Nation, 2001; Hu & Nassaji, 2016). The information related to the three groups is summarized in Table 5.

**Table 4. Information about participants**

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Attribute values</th>
<th>Counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>18-21</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>22-25</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>25-28</td>
<td>1</td>
</tr>
<tr>
<td>Programmes</td>
<td>Business studies</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Engineering</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>Humanities, social sciences</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Biology and medicine</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Physical sciences</td>
<td>25</td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>56</td>
</tr>
<tr>
<td>Region</td>
<td>Hong Kong</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>Mainland</td>
<td>51</td>
</tr>
</tbody>
</table>

**Table 5. Allocation of participants to groups**

<table>
<thead>
<tr>
<th>Groups</th>
<th>Learning strategy</th>
<th>Counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Non-personalized learning (self-paced learning)</td>
<td>35</td>
</tr>
<tr>
<td>B</td>
<td>Personalized learning using the earlier version of the TFA with three components</td>
<td>35</td>
</tr>
<tr>
<td>C</td>
<td>Personalized learning using the current version of the TFA with five components</td>
<td>35</td>
</tr>
</tbody>
</table>

**Experimental procedures**

To verify the effectiveness of the proposed system in promoting vocabulary learning, the following experimental procedures with two stages, as shown in Figure 4, were developed and conducted.

![Figure 4. The experimental procedures](image)

**The first stage**

The first stage involves two steps. Firstly, the subjects were invited to participate in an online training workshop on how the system can be used. After this, the participants’ prior knowledge of 60 candidate target words were evaluated. Based on the test results, 40 words that were unknown to almost every participant were selected as the target words of this study, and thus the participants’ pre-knowledge of the 40 target words was almost zero. All participants were randomly grouped into three teams.

**The second stage**

The second stage also includes two steps: (1) vocabulary learning via the system; and (2) post-testing. The participants were asked to learn the 40 target words through performing different tasks as recommended by the system in a week. The system includes a bank of 20 types of tasks, and the participants were asked to complete
at least one task per day. The three groups of participants applied three approaches to word learning: a non-personalized approach for Group A; a personalized approach guided by a partial version of TFA for Group B; and a personalized approach guided by the full list of TFA for Group C. Specifically, the participants of Group A decided what tasks to do by themselves, as no recommendations were given by the system. Tasks were recommended to the participants of Group B based on the earlier version of the TFA. The participants of Group C were recommended with tasks based on the current version of the TFA. After one week of learning, all participants were tested to evaluate their knowledge of the target words by utilizing a modified vocabulary knowledge scale. This assessment tool has been used in Zou (2017), and the same grading criteria used by Zou (2016) were employed to mark the post-test.

Learning logs

There were 1057 learning logs in total, recording the tasks that were completed by different users during their learning processes. Each user completed a total of 10.07 tasks on average, and 1.43 tasks per day. The maximum number of tasks that were completed by a single user was 20, and the minimum was seven. This is probably because the participants were required to complete one task per day. The distribution of the numbers of learning logs is presented in Figure 5. The vertical axis denotes the number of users, and the horizontal axis denotes the number of logs. 80% of users (i.e., 84 users) completed 7 to 12 tasks within the learning period. There was no significant difference among the numbers of tasks completed by the three groups of participants.

Results

As shown in Table 6, the proposed learning system is very effective in promoting the learning of the target words, given that the participants’ prior knowledge of these words was almost zero. The learning performance of the participants of Group C (a mean score of 74.74) was better than that of the participants in Group B (a mean score of 68.28), and the mean score of the participants in Group A (a score of 61.60) was the lowest among the three groups.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>35</td>
<td>61.60</td>
<td>9.20</td>
</tr>
<tr>
<td>Group B</td>
<td>35</td>
<td>68.28</td>
<td>8.47</td>
</tr>
<tr>
<td>Group C</td>
<td>35</td>
<td>74.74</td>
<td>8.68</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
<th>η²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between groups</td>
<td>3023.16</td>
<td>2</td>
<td>1511.58</td>
<td>19.54</td>
<td>0.00</td>
</tr>
<tr>
<td>Within groups</td>
<td>7888.22</td>
<td>102</td>
<td>77.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>10911.39</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 5. The distribution of the numbers of learning logs
To further examine whether any significant differences existed among the three groups, a one-way ANOVA test was applied, the results of which, as demonstrated in Table 7, indicated statistical significance \((F (2, 102) = 19.54, p < .001, \eta^2 = .28)\).

**Discussion and conclusion**

The results of the research provide empirical supporting evidence for the effectiveness of the proposed e-learning system which recommends word learning tasks based on the TFA scores of different tasks and user models, as the group of participants who learned the target words through a personalized approach guided by the full list of TFA criteria had the best learning performance among the three groups. The effectiveness of this personalized recommendation system results, to a large extent, from the fact that it provides a personalized learning experience and suggests tasks based on TFA utility and task diversity according to users’ learning logs and user models. The results also indicated that e-learning systems should be designed based on comprehensive learning theories, because the group of participants who learned the target words through an approach guided by a partial list of TFA criteria had less effective learning performance than the participants whose learning approach was guided by the full list.

Moreover, it is suggested that effective word learning requires encountering or processing information of target words in different circumstances while performing a wide range of word learning tasks. In other words, a combination of several tasks with different TFA scores is more conducive to word learning than simple repetition of similar tasks. Because tasks with different focuses on the five main components of the TFA checklist promote the learning of different aspects of knowledge of the target words, a combination of different focuses entails a higher probability of building up networks of the target words. Learners are advised to perform tasks with lower TFA scores first, and then move to tasks with higher scores for the learning of certain target words. Additionally, the results further support the argument that a personalized vocabulary learning system is more conducive to word learning than a non-personalized system.

In sum, this study developed a personalized vocabulary learning system under the umbrella of the checklist for TFA, and the experiment results demonstrated that the proposed system is very effective in promoting word learning. The major contribution of the study is to evidence that a comprehensive theoretical framework is essential for the optimal design of learning systems. This research also implies that language education should move toward employing more personalized learning systems, considering that they are very conducive to language learning.

Future studies are suggested to focus on: (1) how to exploit user models to facilitate the learning of various aspects of word knowledge; (2) how to better generate vocabulary learning tasks by using the checklist of the technique feature analysis; and (3) how to integrate the prior knowledge and learning styles of participants in the recommendation process so as to improve the personalized learning experience.

**Acknowledgements**

This study was fully supported by the Start-Up Research Grant (RG 54/2017-2018R) and the Internal Research Grant (RG 63/17-18R) of The Education University of Hong Kong, and the Dean's Reserve-funded Learning and Teaching Project (75.8ACD), The Hong Kong Polytechnic University.

**References**


What Learning Analytics Tells Us: Group Behavior Analysis and Individual Learning Diagnosis based on Long-Term and Large-Scale Data

Jia-Hua Zhang1, Ye-Xing Zhang1, Qin Zou2 and Sen Huang1

1Department of Educational Technology, Zhejiang Normal University, China // 2Dongfang College, Fujian Agriculture Forestry University, China // zjnuzjh@zjnu.edu.cn // zhangyx0216@gmail.com // etzoey@163.com // huangsen0414@gmail.com

Corresponding author

ABSTRACT

The practice and application of education data mining and learning analytics has become the focus of educational researchers. However, it is still a difficult task to explore the law of group learning and the characteristics of individual learning. In this study, the online learning logs of 1,088 students from 22 classes were analyzed from the aspects of their login behaviors, resource utilization, quizzes, interactive behaviors, and academic achievement. To address these issues, multiple methods, including statistical analysis, visualization social network analysis and correlation analysis, were used to analyze the process and results of online learning. The results reveal the characteristics of group behavior of online learners and highlight the key factors that influence the learning process and outcomes of individual learners. From the view of students, these factors include the length and allocation of online time, the effective utilization of resources, social interaction, online learning support and services, etc. From the perspective of teachers, the factors include the management of online teaching, the appropriateness of learning resources, the effectiveness of online intervention strategies, the accurate feedback for online learners, etc. Therefore, learning analysis technology can not only standardize the assessment of learning outcomes, but can also focus more attention on the standardization of learning process assessment. It also identifies the main factors that affect the online learning outcomes and the group characteristics of online learners. At the same time, it provides the learners with personalized learning diagnosis reports which can both help learners understand their own learning status and promote instructors’ accurate teaching and reasonable evaluation.

Keywords

Learning analytics, Learning analysis, Learning diagnostic, Online learning, Educational big data

Introduction

The rapid development of science technology and the popularity of the Internet have promoted the growth and accumulation of data with amazing speed. The arrival of the era of big data indicates a new trend in the development of information technology. Therefore, the concept of big data has been of great interest to all sectors of society since it was first proposed. Of particular interest is the fact that it brings extraordinary innovation and challenges to the development mode and decision-making methods in the fields such as economics, culture, and politics. Education is no exception, with teaching assessment methods, education management tools, learning behavior analysis, and professional development all affected by big data.

The term “Big data” is similar in meaning to “information explosion,” “massive data,” and so on. It is difficult to describe the specific meaning of these words. The McKinsey Global Institute defines big data as a database which is beyond the grasp, storage, management, and analysis capabilities of traditional database software tools. It states: “The massive data mining and application may indicate a new wave of productivity growth and the arrival of the wave of consumer surplus, and data has penetrated into every field” (Manyika et al., 2011). Thus, big data are the data which can be effectively used, and contain huge amount of valuable information and a variety of statistical data.

With the popularization of online teaching and learning, more and more learning platforms are being applied in education. Digital learning information and educational data are growing rapidly. It has become a challenging issue to transform these educational data into useful information and knowledge which affects both students’ learning effect and instructors’ teaching modes as well as the decision-making of educational management and resource allocation. In addition, the new trend of intelligence of learning environments needs deep research on learning analytics. A smart learning environment can offer instant and adaptive support to learners by immediate analysis of the needs of individual learners from different perspectives (Hwang, 2014). Therefore, it has become a hot research topic for many researchers to explore how to use educational data in scientific and effective learning analysis.

However, there is still a lack of long-term and large-scale teaching practice in the field of learning analysis at present. Most of the case studies in this field have used relatively small samples (Martin & Ndoye, 2016; Kim,
Park, Yoon, & Jo, 2016; Sun, Lin, & Chou, 2016; Jin & Kim, 2016; Bainbridge, Melitski, Zahradnik, Lauría, Jayapракash, & Baron, 2015), not to mention the lack of long-term continuous large-scale learning behavior analysis. Therefore, this study reports the use of multiple learning analytics tools and methods in a large-scale and long-term teaching practice. The online learning data of 1,088 students from 22 classes were analyzed to reveal their behavioral characteristics and to provide personalized diagnostic reports to show the value and significance of learning analytics in education settings.

**Literature review**

**Applications of learning analytics**

Learning analytics refers to the use of learners’ data and analysis models to discover information and social connections, and to predict and advise on learning (Siemens, 2010). It is a multidisciplinary method based on data processing technology, educational data mining, and visual data analysis (Scheffel, Drachsler, Stoyanov, & Specht, 2014). To implement continuous formative evaluation and performance evaluation, learning analytics focuses on the measurement of learning and the assessment of academic risk (such as failure to pass the exam, dropping out of the course). Through presenting facts and problems based on learning data, instructors and academic administrators can take measures to intervene when high risk learners are identified, and hence advice can be provided to reduce their academic risk (Pistilli & Arnold, 2010; Lauría, Baron, Devireddy, Sundararaju, & Jayapракash, 2012). Some researchers have used learning data to predict students’ performance, and have provided appropriate interventions for students who were at risk in their learning (Atif, Richards, Bilgin, & Marrone, 2013; Zhang, Fei, Quddus, & Davis, 2014). In addition, some scholars have done research on the use of big data to analyze students’ behaviors in higher education, such as reading materials, submitting homework, interacting with classmates, and taking examinations in online courses. It helped to identify learners with poor performance, and hence to provide suggestions and guidance for improvement (Bienkowski, Feng, & Means, 2012; Lauría et al., 2012; Freitas et al., 2015).

Clow (2013) pointed out that teachers should participate in the process of learning analytics to improve their teaching quality. Using some specific tools, such as learning analytics dashboards, managers, visualization tools, and so on to analyze and monitor students’ online learning behavior can provide students with learning support and improve their academic performance. For example, Zapparolli and Stiubiener (2017) presented a FAG Tool integrated with Moodle to help teachers monitor the participation of their students and take corrective measures in the teaching process. Kim, Jo and Park (2016) indicated that a dashboard can encourage learners, and provide support and service for learners with different academic achievement levels. Echeverría, Benitez, Buendia, Cobos, and Morales (2016) used a Learning Analytics Manager to monitor the processes of collaborative activities and the students’ motivation. Lai and Hwang (2015) utilized a spreadsheet-based visualized mindtool to improve students’ academic performance and their learning attitude. Martin and Ndoye (2016) found that data analysis and the visualization tools were useful in terms of improving teaching and learning. Leeuwen, Janssen, Erkens, and Brekelmans (2015) used two learning analytics tools to increase teachers’ confidence and help students benefit more from the teacher’s presence.

The application of different methods in learning analysis can present different data analysis results for teachers and students. Clow (2013) suggested that learning analytics should link to the management approaches that focus on quantitative metrics. These methods include data statistics and visualization analysis, correlation analysis, social network analysis, clustering and outlier analysis, and so on (Macfadyen & Dawson, 2010; Shiri, 2016). For example, correlation analysis can help the instructor to analyze the relevance between students’ learning behavior and performance, so as to make more scientific decision-making and improve the process of teaching and learning (Sun et al., 2016; Martin & Whitmer, 2016; Auvinen, Hakulinen, & Malmi, 2015; Firat, 2016). Content and sequence analysis can be used to verify the effectiveness of specific learning methods for improving academic performance (Hwang & Chen, 2017; Hwang, Hsu, Lai, & Hsueh, 2017). Teachers and researchers can take specific methods to obtain the corresponding learning analysis results, then provide appropriate services for their students.

Applying some specific tools and methods in learning analysis to exploit data on platforms (such as Blackboard, Moodle) has been a key issue that many researchers have focused on. Whether these tools and methods are appropriately used directly affects the results and value of the data analysis. Moreover, using a predictive model in LMS may help to provide effective intervention and support for at-risk students in a timely manner. For example, Purdue University used a prediction algorithm to determine the risk of students and to provide intervention based on the data collected from the curriculum management system (Arnold & Pistilli, 2012).
Some researchers respectively designed and applied learning analysis and forecasting models to identify those students at risk (Piech, Sahami, Koller, Cooper, & Blikstein, 2012; Kim, Park, Yoon, & Jo, 2016). These studies also indicated that the high precision prediction model provided the possibility of early detection of high risk students and timely intervention.

The challenges faced by learning analytics research

The research of learning analytics has become deeper, partly because of the expansion of data storage capacity and the development of research ability. Due to the convenience of data acquisition and the availability of data, learning analytics can provide a better model for online learning and interaction. At the same time, it also faces many challenges. Avella, Kebritchi, Nunn, and Kanai (2016) pointed out that learning analytics emphasizes the need for more in-depth understanding of how to analyze data and use information from the learning process to optimize results in the education process. In addition, better standardized assessment can improve student participation, while also helping to improve the level of the learner’s ability. Moreover, few studies have focused on how to obtain the data of standard learning activities, which are independent of the teaching methods and learning behaviors in the e-learning environment (Jin & Kim, 2016).

In general, the main challenges faced by learning analytics research are the following: (1) Most teaching activities are still face-to-face in traditional classrooms, and few technologies are used in the learning process. It makes the data of the students’ learning processes difficult to collect. (2) The functions of the tools used for learning analysis are limited, and most data statistics and analysis work is carried out manually. This not only increases the workload and brings inconvenience for the instructors and researchers, but also means that some of the value hidden in the data cannot be fully excavated. (3) The tools and methods used in the learning analysis lack standardization, causing some limitations and inconformity in the research. It is therefore a potential but important issue to propose effective data integration, cleansing methods and management tools for processing educational data (Hwang, Chu, & Yin, 2017). (4) Because of the complexity of educational data collection and the sparsity of educational data structures, the depth of learning data analysis is currently insufficient, meaning that much of the value hidden within the education data cannot be fully excavated. Baker (2016) suggested that learning analytics had considerable potential for achieving high quality adaptive personalization of learning, rather than simply evaluating. (5) The ethical problem in learning analytics also faces challenges. Use of the data collected can infringe on the learners’ personal privacy. In order to solve the ethical dilemma, researchers need to improve the data transparency, and must obtain the consent from learners in the research (Lawson, Beer, Rossi, Moore, & Fleming, 2016).

Research questions

This research used a variety of tools and methods to mine learning data. The data generated by a massive blended course were analyzed from two aspects, namely group level and individual level. The concrete method, process and results of the education data analysis are displayed. It provides personalized learning feedback and diagnostic reports for learners, as well as accurate teaching feedback and decision-making reference for teachers and administrators. In so doing, the study aimed to address the following research questions:

- What are the characteristics of students’ online learning behavior?
- Which data variables in Moodle correlate significantly with students’ achievement?
- How can teachers provide personalized learning feedback for students?

Method

This study is based on an online course on the Moodle platform. The course combined online and offline ways of teaching, and lasted from September 2015 to January 2016. A total of 22 classes and 1,088 learners taking this course participated in the study. Students who take the course need to participate in traditional face-to-face instruction in the classroom, and also log into the Moodle Platform for online learning. Teachers and teaching assistants provide the necessary support services throughout the whole semester.

In this study, multiple methods were adopted, including statistical analysis, visualization, social network analysis and correlation analysis. Some third-party plugins for Moodle (including Gismo, Course Dedication, and Forum Graphs and so on) were used to analyze students’ behavior with the idea of Big Data and educational data mining technology. Through analyzing the behavior of learning groups and individuals, some interesting principles and
characteristics of online learning were revealed. In addition, it provides students, teachers, and administrators with useful information for learning diagnosis and decision-making purposes.

**Results**

**Group behavior analysis of online learning**

Moodle can record detailed information of students’ access to each task module. In this paper, multiple plugins for Moodle were used to collect data according to the research needs. Nine variables were identified to be tracked in this course, which may relate to students’ academic achievement in the course. These variables include number of access times, online time, browsing resource, access frequency, forum posted, forum replied, assignment submission, access period of time, and quiz attempts. The various types of operational behavior of the students involved in this course learning were collected, processed, and analyzed. Finally, the visual analysis results are presented.

**Analysis of access behavior**

Through data collection and behavior statistics, the basic information of students’ access to the online course is shown as Table 1.

<table>
<thead>
<tr>
<th>Total students</th>
<th>Average number of accesses</th>
<th>Maximum number of accesses</th>
<th>Average number of access days</th>
<th>Maximum number of access days</th>
</tr>
</thead>
<tbody>
<tr>
<td>1088</td>
<td>522</td>
<td>52632</td>
<td>12</td>
<td>49</td>
</tr>
</tbody>
</table>

To analyze students’ information of accessing the online learning platform, we take the “week” as a statistical unit to calculate the access rate, which is equal to the quotient of the total access number in the semester divided by the access number in one week. The result is shown as the left part of Figure 1, which indicates that students accessed the online course with a relatively stable frequency in the first 15 weeks, except for the 1st week and 4th week, which both had extremely low frequencies. From the 15th week, the rate began to rise significantly and reached its peak in the 19th week. Thus, the students’ learning behavior was relatively concentrated near the end of the last month of the semester. The access frequency was significantly lower at the beginning of the semester compared with the highest rate at the end of the semester. Therefore, the teacher may intervene with those students who access the online course with very low frequency during the semester. For example, teachers need to urge students to participate in the activities and tasks in the online course in the 5th week.

The analysis result reveals the concentration and dispersion law of the students’ learning behavior in each week. This feature is also reflected in a smaller unit of time, such as each day of a week. We calculated the students’ access rate for a 7-day week as statistical items, as shown in the right part of Figure 1. Students’ participation rates were relatively lower on Thursday, Friday, and Saturday than on the other days. This reflects the characteristics of students’ learning behavior on campus to some extent. It seems that they have more enthusiasm in the first three days of the week, and they may be used to completely reviewing and previewing tasks on weekends.

In addition, the regularity of the students’ behavior varies in the different periods of a whole day. We divided one day into four time periods, which include 0:00-6:00, 6:00-12:00, 12:00-18:00 and 18:00-24:00. The login rates of
these periods are 2.85%, 19.11%, 37.56%, and 40.48%. It indicates that students’ online learning mainly occurs in the afternoon and evening. The peak hours focus on the six hours from 18:00 to 24:00. However, there are still a few students who learn online in the period of 00:00–06:00. They may be a relatively small group of hard-working students. The statistics result shows that 64 learners learned in the early morning, of whom 48 got high scores in the comprehensive performance evaluation. Their scores exceeded more than three-quarters of all learners in the course. Most of these students also stayed up late before the final examination. This may manifest that they paid much attention to the tasks of the online course or just dealt with the final examination. Whatever the reason, this special group of students needs attention.

In order to analyze the relationship between online time input and gender, an independent sample t test was performed. As shown in Table 2, the mean difference of online input time (minutes) between boys and girls was significant. The mean time of girls was 464.45, and that of boys was 252.98, where \( p = .000 < .001 \), with statistically significant contributors. The results showed that there was a significant difference in online input time between boys and girls, with girls spending more time online than boys. Further analysis revealed that the average overall score for girls is 9.18 points higher than that for boys.

Table 2. The independent sample t test on online time input of students of different genders

<table>
<thead>
<tr>
<th>Online time input</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Girls</td>
<td>878</td>
<td>463.45</td>
<td>244.953</td>
<td>11.867***</td>
</tr>
<tr>
<td>Boys</td>
<td>210</td>
<td>252.98</td>
<td>158.855</td>
<td></td>
</tr>
</tbody>
</table>

Note. *** \( p < .001 \).

The blended course requires students to finish different types of online tasks and offline tasks. The correlation analysis was made to verify the correlation between online input time and the scores of different tasks. The result shows that the five parts of the overall scores are significantly correlated with online input time. Table 3 lists the common scores (\( r = .205, p < .01 \)), experiment scores (\( r = .206, p < .01 \)), courseware scores (\( r = .214, p < .01 \)), micro video scores (\( r = .204, p < .01 \)), and online examination scores (\( r = .479, p < .01 \)). It indicates that more online input time will result in higher scores.

Table 3. The relationship between online time input and students’ scores

<table>
<thead>
<tr>
<th>Online time input</th>
<th>Common scores</th>
<th>Experiment scores</th>
<th>Courseware scores</th>
<th>Micro Video scores</th>
<th>Online Exam scores</th>
<th>Overall scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.205**</td>
<td>.206**</td>
<td>.214**</td>
<td>.204**</td>
<td>.479**</td>
<td>.446**</td>
</tr>
</tbody>
</table>

Note. ** \( p < .01 \).

Analysis of resource utilization

The online course is composed of different modules, as shown in Table 4. The page view of each module differs from high to low as follows: quiz, thematic discussion, assignment submission, expanded resource, daily communication, courseware download, and course notification. This order is totally different from that of the layout of each module in the online course. It indicates that the position of resources does not affect the extent of their use.

Table 4. Course description of each module

<table>
<thead>
<tr>
<th>Module</th>
<th>Pageview</th>
<th>Students</th>
<th>Participation rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course notification</td>
<td>969</td>
<td>498</td>
<td>45.77%</td>
</tr>
<tr>
<td>Daily interaction</td>
<td>3455</td>
<td>579</td>
<td>53.22%</td>
</tr>
<tr>
<td>Thematic discussion</td>
<td>42024</td>
<td>1045</td>
<td>96.05%</td>
</tr>
<tr>
<td>Assignment submission</td>
<td>35568</td>
<td>934</td>
<td>85.85%</td>
</tr>
<tr>
<td>Courseware download</td>
<td>3296</td>
<td>670</td>
<td>61.58%</td>
</tr>
<tr>
<td>Expanded resources</td>
<td>5299</td>
<td>719</td>
<td>66.08%</td>
</tr>
<tr>
<td>Quizzes</td>
<td>303941</td>
<td>1011</td>
<td>92.92%</td>
</tr>
</tbody>
</table>

The curriculum teaching content of “Modern Educational technology” contains 11 themes (chapters). The visit of each chapter is shown in Figure 2. As an online learning resource, the degree of utilization of each resource and the importance of the resources to the different students differ greatly. Students’ participation is the highest for Topic0 and Topic1, while the participation for the following subjects is significantly lower. The main reason for this phenomenon is that Topic0 and Topic1 are directly related to the classroom learning tasks or experiment tasks, whereas the other resources are not related.
From the learner’s access to the resources of individual cases, the individual behavior of different learners is also very obvious. For analysis purposes, they were divided into a positive group (Access ≥ 25) and a negative group (Access ≤ 5). The average access of the positive group was about 5 times that of the negative group, as shown in Figure 3. Combined with the final scores of the analysis, the resource access is proportional to academic performance.

In addition, there are nine teachers who taught 22 classes in this research. Each teacher taught a different number of classes. Utilization of learning resources in different classes is shown in Figure 4. It was found that there were 11 classes with the highest average score, in which the students were mainly majoring in Chinese language and literature, primary education, humanities education, and English. These classes were taught by six teachers. The average resource utilization rate was close to 60% in these 11 classes. Learning resources are therefore an important influencing factor of online learning.

Analysis of the quizzes

In order to allow students to strengthen their understanding of the knowledge and to consolidate their knowledge in time, the online course provides several quizzes according to the content of each chapter. Students can receive timely feedback and make up the shortfall through quizzes and repeated exercises. We can see the completion
rate of all quizzes from Figure 5. Because this module is closely related to the final examination, its overall completion rate is higher and better than that of other modules.

![Figure 5. Completion of 10 quizzes](image)

**Interactive behavior analysis**

From the time distribution of the discussion and daily communication of the students, the same characteristic of the learning time is presented. As shown in Figure 6, it is mainly concentrated in the last month, when the frequency of discussion significantly increases. Through the analysis of the interactive relationship between teachers and students with social networks, we found diverse and interesting modes of interaction among students and teachers: (1) Single-center mode. It is usually found in the required discussion area in this course. All the members are in the same social network. There are no isolated nodes. A poster (teacher) is located in the center of the discussion. Other members post around the central node, but interaction between different members is rare, as shown in Figure 7-A. (2) Multi-center mode. Different classes of teachers and students carry out free discussions based on different topics in daily communication. They maintain a high degree of enthusiasm and participation that shows frequent interaction. The more active members have a greater number of nodes in the graph and a closer connection, as shown in Figure 7-B. (3) Small-group mode. Some students participate in an optional discussion area, and active individuals become the opinion leaders. Around these leaders, several small groups are set up, as shown in Figure 7-C. (4) Isolated-individual mode. In some optional topics, few students participated. They only expressed their opinion, but did not care about others’ opinions. In this mode, each node is isolated from others, as shown in Figure 7-D.

![Figure 6. Forum access time distribution](image)

**Academic achievement analysis**

Numerous factors influence academic achievement. However, many students are often only concerned about the modules which are directly related to their final grades, and easily ignore some implicit modules. From the learning process evaluation and comprehensive evaluation, the final grade mainly includes normal performance, assignment performance (such as experiment, micro-video, courseware) and the final examination. According to the level of student achievement, the students are divided into three groups: excellent, good and general. More than half of the students got good scores in this course. It is worth noting that the teachers did not set a mandatory requirement for access to the learning resources. However, 56.75% of the students persisted in online learning throughout the whole semester. The extent of students paying attention in different modules is closely related to the requirement of learning tasks and assessment. In addition, most of the students who got the highest page views scored 90% or more. On the contrary, the students with the lowest page views often scored much lower.
lower. In some classes, we can find a significant positive correlation between the resource utilization rate and average score. This means that learning resources are an important factor affecting student achievement, but not the absolute and only factor.

In addition, the average grade differed by teacher and class. Even different classes with the same teacher showed significant differences, which is shown in Figure 8. For example, four classes were taught by teacher F, and the total academic achievement of these classes ranked in the top 10 of all classes. They achieved better results because the average resource utilization rate of the four classes was 64.18%, which is higher than the average resource utilization rate of 56.75%. In addition, Teacher F emphasized interaction with the students on the online platform. Teachers and students initiated and replied to more than 400 posts, and each student posted and replied at least once in the discussion area. The teacher set up a good learning atmosphere and promoted students’ enthusiasm. Outstanding comprehensive performance is also attributed to the average of more than 85 points of normal performance, assignment performance, experimental performance, and the final grade. Most notably, the normal performance and experiment performance are both more than 90 points. In other words, online learning attendance is affected by many factors, such as teaching methods, learning resources, learning activities, basics of learners, learning atmosphere, etc. However, the most important factor is students’ enthusiasm and initiative for online learning tasks and activities.

The correlation analysis was carried out to further verify the relationship between the overall score and resource access, quiz attempts, and forum browsing, as shown in Table 5. When the confidence (double sides) between
the overall scores and the resource browsing is 0.01, the correlation is obvious. That is, based on the whole tendency, the higher the students’ overall scores, the more the amount of resource browsing. The correlation between overall scores and quiz attempts is that when the confidence (double sides) is 0.01, the correlation is obvious, which means that the higher the students’ overall scores, the higher the quiz completion rate. The correlation between students’ overall scores and forum browsing is obvious when the confidence (double sides) between the overall scores and the forum browsing is 0.05. That is, the higher the students’ overall scores, the more the amount of forum browsing, but the correlation is a little weaker.

<p>| Table 5. Correlation analysis between students’ overall scores and the key online behaviors (N = 1088) |
|---------------------------------------------------|----------------|----------------|----------------|</p>
<table>
<thead>
<tr>
<th>Students’ overall scores</th>
<th>Resource access</th>
<th>Quiz attempts</th>
<th>Forum browsing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.470**</td>
<td>.362**</td>
<td>.074</td>
</tr>
</tbody>
</table>

Note. * p < .05; ** p < .01.

Individual diagnosis of online learning

By analyzing the overall learning situation, we can provide personalized diagnostic reports for each student on the basis of the existing data. These reports include the learning process, outcomes, problems, and recommendation through visual presentation for the whole semester. Visual comparison of the results also includes personal and class average grades, so that students can clearly understand their own performance. What’s more, they are provided with evidence of learning reflection on the learning platform. The teachers therefore have a comprehensive and detailed understanding of the students’ learning status in a data-driven assessment. It can also provide valuable reference for improving teaching methods, curriculum construction, learning resource development, learning activities design, supervision, and intervention of the learning process. At the same time, it provides useful information to optimize the allocation of resources for academic administrators. As shown in Figure 9, a student was selected as a sample to show his personal learning diagnosis report. It provides a clear description of the student online learning process and outcome, which can help the student to know the problems in his/her online learning behavior.

As shown in Figure 9, the student’s learning diagnosis report is composed of six aspects:

- **Access Frequency Analysis.** This module is mainly to compare the access frequency of individual students and the average access frequency of all students in the learning platform. As shown in the figure, the student’s total access number is 382 times, which is lower than the average access number of 522 times. Therefore, the student was not so active in the learning process during the semester.

- **Resource utilization rate.** This module is mainly to compare the resource utilization rate of individual students and the average utilization rate of all the students in the learning platform. As shown in the figure, the students’ utilization rate is only 1.27%, which is far below the average utilization rate of 61.38%. This indicates that the student did not make full use of the learning resources provided in the online course. The student’s attention to the learning resources needs to be improved.

- **Completion of quizzes.** This module is mainly to compare the students’ completion rate of the quizzes and the average rate of all the students in the learning platform. As shown in the figure, this student’s completion rate for the quizzes is relatively poor. In some chapters, he failed to pass the quizzes, and his average score is about 20 points lower than the grade and class average. The student therefore needs to practice more.

- **Forum interaction.** This module is mainly to compare the interaction between the students and the grade in the forum. As shown in the figure, the student’s interactions are as follows in the forum: total read times of 39 and reply times of 3. The average read times is 42 and the reply times is 3. The student is thus active in the forum.

- **Score of all kinds of achievements.** This module is mainly to compare all kinds of individual scores and average scores of the class and grade. As shown in Figure 9, the final examination score is about 30 points lower than the average for the grade and the class in all kinds of achievement. However, his usual performance exceeds the grade average.

- **The distribution of comprehensive performance.** This module is mainly to compare a student’s score with the comprehensive results of the whole grade and class. For example, the student’s comprehensive score is 75, which indicates a general level, but it is lower than the average score of the whole grade and class.
Finally, the individual learning diagnosis report shows a detailed understanding of the student’s learning situation. Some suggestions are provided for the students from all aspects, which is shown in Figure 9. The individual learning diagnosis report is a visual representation of the state and outcomes of learning. It transforms data into valuable information and helps learners to understand their own learning participation in all dimensions and multiple perspectives, which forms a digital self-analysis and summary.
Discussion and conclusions

Characteristics of online learning

Learning analysis can help teachers understand the characteristics of students’ online learning behaviors and provide the basis for the adjustment and optimization of follow-up teaching. In this study, in terms of learning time, students’ participation frequency grew from slow to fast. At the beginning and mid-term, the attendance was relatively low, but there was a rapid rise at the end of the semester. This indicates that the students’ online learning time allocation was unbalanced and unreasonable. In this course, the teachers had no rigid demands regarding students’ study time, so the students did not spend much time on the learning platform, especially at the beginning and in the middle of the course. However, they paid more attention to the learning resources and activities which were closely related to the final examination, which led to a sharp increase in student participation at the end of the semester. Therefore, the teachers should actively advocate that the students participate in online learning. If necessary, teachers can also forcibly set the minimum online learning time for students to increase their participation in the course.

From the aspect of learning resources, the resource utilization rate is relatively low. The main reason is that no corresponding intervention has been implemented on students. Therefore, the frequency of learners accessing online resources varies irregularly. For example, the frequency of accessing PowerPoint slides increased dramatically before the end of term. This may indicate that students need to review the content before they finish the quizzes of each chapter. The teachers therefore need to optimize the presentation of resources and enhance the attractiveness of these resources. Also, teachers can establish appropriate monitoring and intervention mechanisms to urge students to effectively use the resources of the online course.

In addition, the Forum Graph plug-in presented the visual results of the online interaction. It indicates that teachers usually act as leaders and serve an important intermediary role in most thematic forums. Most learners can communicate with their classmates and teachers. However, only a few students initiated discussion and exchanged ideas with others. For most students, their interaction in optional tasks were particularly weak. The possible reason is that the participation of teachers in the forum was not high. They were only the initiators of topics and did not pay enough attention to students’ performance in the forum. Hernández-García, González-González, Jiménez-Zarco, and Chaparro-Peláe (2015) found that a lack of instructor’s activity couldn’t ensure better group performance, and the students who got more replies from teachers tended to get higher grades. Therefore, teachers should actively participate in the discussion and pay more attention to students’ discussion contents, then give them the appropriate answer in time. In addition, the teachers may select more interesting topics and try active and reasonable strategies to stimulate students’ participation.

Significant indicators of learning achievement

Learning analysis can identify important indicators that affect academic performance, then help teachers make more scientific decisions. In this study, the students’ resource utilization, quiz attempts, forum browsing, and overall scores were correlation analyzed by Pearson analysis using the SPSS software. The conclusions are as follows. Students’ resource access, quiz attempts and forum browsing showed positive correlations with their overall scores. Among them, the correlation between resource access, quiz attempts, and overall scores was strong, but the correlation between forum browsing and overall scores was relatively weak. This conclusion is somewhat different from that of Bainbridge et al. (2015) who showed that relative number of forum posts and amount of content read were the most important predictors of success in a course. In this study, the reason why the relevance of the forum browsing and overall scores was weak has two aspects. On one hand, it may be due to the fact that the proportion of the number of forum posts in the overall score was small, making students not pay enough attention. On the other hand, it may be due to the lack of interaction between teachers and students, also resulting in weak interaction among students. The four models of forum interaction presented above also predicted this result to a certain extent.

Therefore, optimization of the assessment method and increase in teacher-student interaction are both important measures to improve online learning behavior. Teachers need to establish a comprehensive assessment mechanism, such as uploading and sharing resources. The topics and the influences of forum interaction should also be evaluated, and more attention should be paid to the interaction between teachers and students. Teachers should also promptly reply and appropriately guide students to discuss the topic, urging them with timely feedback. If teachers increase their interaction with the students, it might not only improve the group
performance, but could also improve the students’ academic achievement. These discoveries above are worthy of teachers’ reflection.

**Learning diagnostic reports for individual students**

The statistical analysis of student group behavior by using learning analysis is aimed at showing the overall situation of students’ online learning. It is helpful for teachers to make the necessary adjustments to their teaching according to the statistical results and to make corresponding demands on students’ online learning to improve their learning process. The analysis of group behavior mainly provides the basis for teachers to improve their teaching. The individual learning diagnosis report can provide teachers with a better understanding of individual students’ learning situations, then they can provide personalized learning intervention for those students who are at risk. It also provides students with personalized learning feedback, so as to help them obtain personal learning experience. For example, students can know the gap between the whole class and themselves on the learning situation through the diagnostic report, so as to adjust learning and improve academic performance according to the learning suggestions. Some researchers also put forward similar visual online behavior models for learners to improve their awareness of the learning activities, which help to improve their academic achievement (Charleer, Klerkx, Duval, Laet, & Verbert, 2016; Ramos-Soto, Vázquez-Barreiros, Bugarín, Gewerc, & Barro, 2017). Therefore, a personalized learning diagnostic report for individual students is necessary.

**Suggestions to teachers**

Teachers should provide diverse and appropriate learning resources. Because of the autonomous characteristic of online learning, the important influencing factors of the academic success of the learners are the learning resources. In order to improve the utilization of learning resources, teachers need to enhance the attractiveness of learning.

Teachers should design appropriate learning and discussion topics. Because online learning time is not fixed and non-realistic, most of the interaction is asynchronous. Teachers need to design discussion topics or encourage learners to issue a discussion topic carefully. In order to attract learners to actively participate in the discussion, good interaction between teachers and students is needed to enhance mutual understanding.

Teachers should publish timely and accurate online learning feedback. Learning feedback enables learners to make clear the current learning progress and learning situation. This is an important basis for adjusting learning methods and learning time in a timely manner. Based on learning analytics, teachers can provide visual feedback of different stages of the learning process to learners, which can facilitate improvement in their learning.

Teachers should adopt effective online learning intervention strategies. Learning analytics can visualize the learning process and the results of the data. According to the results, teachers can be provided with quantitative, objective, and timely feedback. It is possible to predict when students are at risk in their academic studies and to take the necessary learning interventions to reduce their academic risk.

**Acknowledgements**

This study is supported by the MOE (Ministry of Education in China) Youth Fund Project of Humanities and Social Sciences (Project No. 17YJC880125).

**References**


A Comparison between Two Automatic Assessment Approaches for Programming: An Empirical Study on MOOCs

**Anis Bey**¹, **Patrick Jermann**² and **Pierre Dillenbourg**³

¹Ecole Supérieure des Sciences de Gestion, LRI, Université Badji Mokhtar-Annaba, Algeria // ²Center for Digital Education, Ecole Polytechnique Fédérale de Lausanne, Switzerland // ³Computer Human Interaction in Learning and Instruction, Ecole Polytechnique Fédérale de Lausanne, Switzerland // anis.bey@univ-annaba.org // patrick.jermann@epfl.ch // pierre.dillenbourg@epfl.ch

³Corresponding author

**ABSTRACT**

Computer-graders have been in regular use in the context of MOOCs (Massive Open Online Courses). The automatic grading of programs presents an opportunity to assess and provide tailored feedback to large classes, while featuring at the same time a number of benefits like: immediate feedback, unlimited submissions, as well as low cost of feedback. The present paper compares Algo+, an automatic assessment tool for computer programs, to an automatic grader used in a MOOC course at EPFL (Ecole Polytechnique Fédérale de Lausanne, Switzerland). This empirical study explores the practicability and behavior of Algo+ and analyzes whether it can be used to evaluate a large scale of programs. Algo+ is a prototype based on a static analysis approach for automated assessment of algorithms where programs are not executed but analyzed by looking at their instructions. The second tool, EPFL grader, is used to grade programs submitted by students in MOOCs of Introductory programming with C++ at EPFL and is based on a compiler approach (Dynamic Analysis approach). In this technique submissions are assessed via a battery of unit tests where the student programs are run with standard input and assessed on whether they produced the correct output. In this study results showed the advantages and limits of each approach and pointed out how the two tools can be used to get a benefit assessment of students’ learning in MOOCs of computer programming. This study led to the proposition of a model for the relationship between the number of submissions and the appearance of the most frequent submitted programs. This technique is used by Algo+ for giving feedback and it is based only on the n most redundant submissions that have been annotated by the instructor.

**Keywords**

Computer education, Programming assignments, MOOCs, Assessment, Automated grading, Static analysis, Dynamic analysis, CS1

**Introduction**

In introductory programming classes, teachers are often faced with grading a huge number of assignments. Because grading requires dealing with the complexities of the submissions, it becomes a strenuous and burdening task for teachers. To deal with this problem, there are generally two ways: increasing the number of teachers and the use of computers. The first approach is often not feasible (because of personnel costs) whereas using computers to run automatic assessment tools allows dealing with this issue. This approach has many advantages: objective assessment, removal of halo effects, and reduction in manual, error-prone and marking burden (Kalogeropoulos et al., 2011).

Computer based assessment is useful for handling very large numbers of students. Apart from giving a score, these tools may also offer an environment to practice programming, which is useful for students to develop programming skills. Since programming cannot be learned solely from books as in other subjects, students have to learn programming by developing algorithms themselves to deepen their understanding (Lahtinen et al., 2005).

In the last years, MOOCs (Massive Open Online Courses) represent the new form of learning and the new way to teach a large number of students without a place or time restriction. A massive number of students subscribe to these courses through several online platforms like EdX, Coursera and Udacity that give multiple opportunities to students to resubmit their code to improve on their mistakes, providing the opportunity for mastery learning. Since assessment resources are not restricted, automated grading allows students to increase mastery by iteratively improving and resubmitting their homework.

A big challenge that needs to be addressed for MOOCs to work is automated assessment of student work. The tremendous number of submissions that require feedback justifies the need for automated assessment tools. Given the popularity and widely claimed promise of MOOCs technology, we submit that how students are assessed and which forms this assessment takes are vital research questions. The involvement of a service for
Another area of special interest is learning programming. In the last decades, programming is known as the 21st century skills. As programming skills become ever more important and a core competency for all kinds of 21st Century jobs, this is leading researchers to seek out new ways of learning to program. But learning and teaching the basics of programming is a complicated task and the best way for learning programming is practicing programming techniques and concepts by trying them out yourself (Lahtinen et al., 2005). Programming is not an exact science, but the more you practice, the more you develop skills; “practice makes perfect,” and to do, students need an efficient assessment tool to tell them if they gave a correct or a wrong solution, and to give an efficient and timely feedback with a summative assessment to quantify achievement.

In the current study, we examined two automated graders: (1) Algo+, a prototype system developed to assess algorithmic competencies (Bey & Bensebaa, 2011; Bey & Bensebaa, 2013), based on static program analysis approach and (2) EPFL grader, an automated grader developed at EPFL (Ecole Polytechnique Fédérale de Lausanne, Switzerland) and based on dynamic program analysis. This tool is used to assess students’ programs in the MOOC Introduction to Programming (with C++). And the question that guided this study was: Could Algo+ tool also be used to automatically assess the submissions of students in the context of MOOCs? To form a better view of the effectiveness of Algo+, we compared its results to those of EPFL grader real data and we examined the behavior of Algo+ tool in the case where it would be used in the context of MOOCs learning.

The rest of this paper is organized as follows. Primarily, we show the different forms of assessment used in MOOCs of computer programming, some related works and the main approaches used in automated assessment in programming. Also, we briefly state the functionalities of Algo+ and EPFL grader. The method design is presented with a description of data collected, results presentation and analysis of both scoring results and types of feedback provided by each tool. Finally, we conclude this study by proposing a model about the relationship between the number of submissions and the number of the appearance of the most frequent submissions.

Assessment in MOOCs of computer programming

Assessment covers an important part of MOOCs and constitutes the key driver for learning. The actual providers of MOOCs like Coursera and edX propose simple tools in the context of programming such as Multiple-Choice Questions and Peer Assessment. How to design assessments is a challenge in itself as MOOCs have massive and diverse student enrollment.

We can usually distinguish three manners to assess computer programming skills in MOOCs. The first uses simply the online assessment through multiple choice questions or multiple choice cloze tests. This kind of assessment is easy to be automated but it is not suitable for assessing conceptual skills in computer programming.

The second form is the Peer Assessment. Self and peer assessment might offer promising solutions that can scale the grading of complex assignments in courses with thousands of students. This form of assessment has been historically used for logistical, pedagogical, metacognitive, and affective benefits. For example, Coursera uses peer grading by submitting an assignment and five people grade it; in turn, you grade five assignments. In general, students becoming assessors may bring multiple spin-offs. It can be a significant learning experience if it was well explained and designed to the students and on the other hand, a huge number of submissions can be graded. But leaving assessment in the hands of the students can generate a depreciated assessment because they may create their own criteria that may not be well founded.

The third way is using a third-party grader tool option. The use of automatic graders is not typical in most of MOOCs providers in their programming courses. They let MOOC’s owners use external graders to provide interactive, dynamic, online coding exercises and more complex programming assignments for learners. The execution of codes must be totally independent as illustrated in Figure 1 to avoid that malicious codes or badly written codes affect the MOOC system.
In general, the choice of technique for the assessment in MOOCs for programming depends on what instructors want to assess in students after following the MOOC and what they want to know about their skills. But it is essential to mention that the assessment process has to take into account some features like assessing problems having different answers or partially correct answers options.

Figure 1. Submission page of a MOOC of programming using a submitting space for writing code

Related works

Automated assessment tools in programming has often been studied since 1960 by Hollingsworth (Hollingsworth, 1960) and are growing in popularity with the appearance of MOOCs to deal with the colossal number of submissions.

We do not provide an exhaustive review of the existent automated grading systems in this paper. Various studies have investigated the value of automated assessment tools and surveys. Ala-Mutka (2005) and Ihantola et al. (2010) listed a comprehensive overview of those systems. We can cite some systems and tools that can be found in almost all literature review studies on automated assessment in computer programming: CAP (Schorsch, 1995), Ceilidh (Benford et al., 1995), ASSYST (Jackson, 2000), CourseMarker (Higgins et al., 2003), TRAKLA2 (Korhonen et al., 2003), Curator (Edwards, 2003), Submit (Harris et al., 2004), TRAKLA (Malmi et al., 2005), BOSS (Joy et al., 2005), Mooshak (Guerreiro & Georgouli, 2006), Autograder (Nordquist, 2007), WebCAT (Edwards & Perez-Quinones, 2008), Athene (Towell & Reeves, 2009), AutoLep (Wang et al., 2011), JUG (Brown et al., 2012), Aari (Taherkhani et al., 2012), COALA (Jurado et al., 2012), CodeWrite (Denny et al., 2014).

WebCAT (Edwards & Perez-Quinones, 2008) is one of the most popular automated assessment tool used by many institutes to assess students’ programs source code. This tool provides some useful features (i) it allows submissions directly from within Eclipse using a simple plug-in (also from a few other IDEs, or directly via a web browser), (ii) instructors can use JUnit tests to check student solutions; Checkstyle and PMD to perform static analysis checks on student solutions, (iii) allows students to write and submit their own JUnit tests. Of course, WebCAT is based only on Java programming language.

To date, there are two main approaches used for grading students’ programs. The first approach is dynamic analysis. It runs a program through a data set and then compares the output to the predefined answer. The second approach does not run student programs but it looks for metrics; such as line of code, number of variables, statements and expressions, or attempts to compare the source code of the student program with a model program.

Although many automated programming assessment systems have been proved to be of great help to both instructors and students in learning programming but several problems remain unsolved. Foremost among these problems are security, algorithms for automatic generation of test data in dynamic analysis, low accuracy and
precision of correctness and functionality assessment in static analysis. However, assessing the functionality of students’ code is still the most often used approach to grade programs due to its simplicity of implementation by running programs.

**Algo+ and EPFL grader overviews**

**EPFL grader**

The EPFL grader is an automated grader used to assess students’ assignments in programming during the MOOCs course Introduction to programming with C++ (a MOOC presented by Jamila Sam, Jean-Cédric Chappelier and Vincent Lepetit through Coursera platform in Ecole Polytechnique Fédérale de Lausanne, Switzerland). For every session, this course provides a set of programming exercises for participants. Exercises can be solved by submitting a source code as a file within the web browser (see Figure 2). Solutions submitted by students are compiled and unit-tested over a set of inputs. In return, students receive a score and an automatic feedback on how their code performed in the tests. There is no limit to submitting solution programs.

![Figure 2. Submission page of Introduction to programming with C++ MOOC at EPFL using the EPFL grader](image)

As illustrated in Figure 3, the automated grader of EPFL programming MOOC is based on two main components: (1) a battery of unit tests where students’ programs are run with standard input and assessed on whether they produced the correct output, and (2) a style checker which may deduct points for bad style in working programs and give feedback. Students are able to experiment with the tool and find out for themselves whether something works or not within a feedback.

**Algo+**

Algo+ is an automated assessment tool of programs using program matching (Bey & Bensebaa, 2011; Bey & Bensebaa, 2013). A submitted program is assessed by comparing it to a set of predefined solutions already assessed by an instructor. Predefined solutions are called Referent Solutions and are those common and frequent
submissions that have been detected in students’ submissions. The set of referent solutions used to assess submitted programs contains not only correct programs but also erroneous ones. The whole solutions (correct and erroneous) are gathered and used as a source of learning and assessment for the new submitted cases (see Figure 4).

![Diagram](image)

**Figure 4.** Algo+ mechanism

Each referent solution in the database has a score and a feedback giving by the instructor. The feedback giving by the instructor for each referent solution should be general. It explains misconceptions, if the case of erroneous solution, or general explanation about how the solution is made if it is correct.

**Algorithm 1:** Algo+ algorithm for assessing students’ programs

```plaintext
Input : RS[nb.RS]; /* list of Referent Solutions
RE[nb.RE]; /* list of Referent Errors
Submission;

Output: Score, feedback1, feedback2
1 Simmax ← Dist(Submission, RS[0]);
2 Ref ← RS[0];
3 for i ← 1 to nb.RS do
4 Sim ← Dist(Submission, RS[i]);
if Simmax < Sim then
5 Simmax ← Sim;
6 Ref ← RS[i];
end
9 end
10 if Simmax = 1 then
11 Score ← Ref.Score;
feedback1 ← Ref.feedback;
else
14 Score ← Ref.Score * Simmax;
15 Simmax ← Dist(Submission, RE[0]);
16 Ref ← RE[0];
17 for i ← 1 to nb.RE do
18 Sim ← Dist(Submission, RE[i]);
if Simmax < Sim then
19 Simmax ← Sim;
20 Ref ← RE[i];
end
22 end
23 end
24 if Simmax > 0.5 then
25 feedback2 ← Ref.feedback;
end
27 end
28 write(Score, feedback1, feedback2);
```

**Algorithm 1:** Algo+’s algorithm describing steps for grading students’ programs
Figure 4 illustrates how a submitted program is assessed. The recognizer module is based on code-matching and it described in the algorithm (see Algorithm 1). Correct referent solutions are used for scoring while erroneous solutions are used to giving feedback. It tries to recognize the submitted program among referent solutions. If the submitted program was recognized among referent solutions, the score and the feedback of the referent solution are returned to the student. Otherwise, the submitted program is considered as unrecognized and stored in frequent submission database. If it appeared more than two times from different students then it would be considered as common and sent to the instructor which has to award an appropriate score and feedback. The score attributed to student in this case is calculated from the similarity value.

Study design

This empirical study was undertaken on real datasets. The comparison was performed with large samples of programs collected from a MOOC course in Introduction to programming with C++ at EPFL. It is one of the important MOOCs courses provided by EPFL. Approximately, 13500 participants follow this MOOC.

The research question asked in this study is: How well could Algo+ assess students’ submissions in a MOOC of programming compared to EPFL grader?

Data collection

In this work, we use datasets collected from a course of Introduction to programming with C++ taught by EPFL professors. All collected submissions were submitted between September 2013 and September 2014. In this course, a grading system is used to assess students' submissions. Students are asked to download a file containing a pre-established source code where it is indicated to put the asked algorithm according to the statement of the exercise.

Two different exercises were selected for this study. The first exercise (Exercise 1) was about swapping values. Students were required to write a C++ program that swap three values. For example, for these input data a = 51, b = 876 and c = 235, we obtain a = 235, b = 51 and c = 876.

The second exercise (Exercise 2) was to write a C++ program that guesses which character (among a list known in advance) the user has in mind. The purpose of this exercise is the ability to use conditional structure.

Fortunately, an electronic archive of all submissions made during the session within the assessment results provided by EPFL grader (submission ID, mark and date) are available. A total of 2312 participants completed Exercise 1 and produced 4985 submissions. Only 313 (n = 3130) programs are used in this study after eliminating programs that are damaged or considered not valuable. It is the case when the code was inserted into the inappropriate place where it has been indicated in the pre-established source code. Otherwise, the submitted program will not be assessed and a value of 0 is automatically attributed. In the second exercise 1991 participants produced 7067 programs and solely 4914 programs (n = 4914) were considered in this experiment.

On Algo+ side we need at least one referent solution to be able to assess submitted programs. In this study, we have started with two referent solutions for each exercise. Scores assigned by the two tools are rated on [0-30] for the first exercise and [0-50] for the second exercise.

Results

This study was designed to determine whether Algo+ can assess students' submissions in MOOCs as well as EPFL grader. The obtained results were reported according to three important features that an automated assessment tool has to have in the context of MOOCs: correctness (summative), usefulness of the feedback (formative) and the speed of assessment (massive).

The main summary descriptive statistics are presented in Table 1. Both measure of center that is mean and median of EPFL grader and Algo+ are almost similar in the first exercise but different in the second exercise. The measure of variation that is standard deviation is also similar in the first exercise but different in the second. The means and medians of Algo+ and EPFL grader scores were (M = 20.26, SD = 14.03) and (M = 21.6, SD =
12.06) in Exercise 1 and \((M = 24.76, SD = 22.2)\) and \((M = 28.96, SD = 17.02)\) in Exercise 2. In addition, the Wilcoxon signed rank test was performed to examine the range of score frequencies and showed that there is a significant difference in the range of score frequencies \((p < .05)\).

| Table 1. Descriptive statistics |
|-----------------|-----------------|-----------------|-----------------|
|                 | Exercise 1      |                 | Exercise 2      |
|                 | Algo+           | EPFL grader     | Algo+           | EPFL grader     |
| Min             | 0.00            | 0.00            | 0.00            | 0.00            |
| 1st quartile    | 0.00            | 10              | 0.00            | 16.35           |
| Median          | 30.00           | 30.00           | 30              | 29.81           |
| Mean            | 20.26           | 21.6            | 24.76           | 28.96           |
| 3rd quartile    | 30.00           | 30.00           | 50              | 50              |
| Max.            | 30.00           | 30.00           | 50              | 50              |
| SD              | 14.03           | 12.06           | 22.20           | 17.02           |
| Mode            | 30              | 30              | 0               | 50              |

Results of correlational analyses, using the nonparametric Spearman Rank Correlation Coefficient tests, indicated that statistically significant correlations were present between Algo+ and EPFL grader in both exercises \((Exercise 1: rs = 0.92, p < .001 and Exercise 2: rs = 0.59, p < .001)\). These results show excellent correspondence between the two sets of marks for both the direct comparison with EPFL grader and with the ranked order of student programs in the two exercises.

To check if there are differences in scores distributions between Algo+ and EPFL grader marks, we have presented in Figure 5 histograms of scores of both Algo+ and EPFL grader in the two exercises. We compared scores distribution using two-sample Kolmogorov-Smirnov tests, which found significant differences for the two exercises \((p < .05)\). In the two exercises in question, we can easily distinguish that scores assigned by EPFL grader were even more concentrated on two values in the first exercise (the minimum score 0 and the largest score 30) and six values in the second exercise (0;10;20;30;40;50). Contrary to Algo+ where scores distribution is concentrated on values ranged between 0-30 and 0-50 for the first and the second exercise, respectively.

![Scores Distribution of Algo+ (Exercise 1)](image1)

![Scores Distribution of EPFL grader (Exercise 1)](image2)

![Scores Distribution of Algo+ (Exercise 2)](image3)

![Scores Distribution of EPFL grader (Exercise 2)](image4)

*Figure 5. Histograms showing scores distribution of Algo+ and EPFL grader*
As scores assigned by Algo+ and EPFL grader are rational numbers (e.g., 2.3; 2.4; 2.5) it is difficult and not very meaningful to calculate kappa or the percentage of absolute agreement. Under these conditions, the inter-rater agreement between Algo+ and EPFL grader was calculated with the intraclass correlation coefficient (ICC) (McGraw & Wong, 1996) with a two-way mixed model using R version 3.2.4. The intraclass correlation (ICC) is a measure of agreement and it is useful when there are many rating categories (5 or more) or when ratings are made along a continuous scale. The “agreement” ICC is the ratio of the subject variance by the sum of the subject variance, the rater variance and the residual. The inter-rater agreement between Algo+ and EPFL grader was substantial in the first exercise (ICC = 0.88, 95% CI 0.86-0.89, p < .001) but it was a moderate agreement in the second exercise (ICC = 0.51, 95% CI 0.48-0.55, p < .001) based on the cut-off value for acceptability level (Graham et al., 2012).

To further investigate scores assigned by the two tools and understand differences, frequencies were performed to examine the distribution of scores. And for ease of comparison and discussion, we presented the frequency tables by three score ranges, namely, by erroneous programs (scores = 0), somewhat correct programs (scores between 0 and the largest score) and correct programs (scores equal to the largest score).

Table 2 and Table 3 present data on similarities and differences in scores between EPFL grader and Algo+ according to these three score ranges (correct, somewhat correct and erroneous).

**Table 2.** Concordance and discordance between Algo+ and EPFL grader in Exercise 1 (n = 3130)

<table>
<thead>
<tr>
<th></th>
<th>Algo+</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correct</td>
<td>Somewhat-Correct</td>
<td>Erroneous</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td>EPFL grader</td>
<td>Correct</td>
<td>2022</td>
<td>38 (A)</td>
<td>39 (B)</td>
<td>2099</td>
</tr>
<tr>
<td></td>
<td>Somewhat-Correct</td>
<td>0</td>
<td>14</td>
<td>5 (C)</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Erroneous</td>
<td>0</td>
<td>557 (D)</td>
<td>455</td>
<td>1012</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>2022</td>
<td>609</td>
<td>499</td>
<td>3130</td>
</tr>
</tbody>
</table>

**Table 3.** Concordance and discordance between Algo+ and EPFL grader in Exercise 2 (n = 4914)

<table>
<thead>
<tr>
<th></th>
<th>Algo+</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correct</td>
<td>1242</td>
<td>49 (A)</td>
<td>362 (B)</td>
<td>1653</td>
</tr>
<tr>
<td></td>
<td>Somewhat-Correct</td>
<td>0</td>
<td>209</td>
<td>1696 (C)</td>
<td>1905</td>
</tr>
<tr>
<td></td>
<td>Erroneous</td>
<td>0</td>
<td>136 (D)</td>
<td>1220</td>
<td>1356</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>1242</td>
<td>394</td>
<td>3278</td>
<td>4914</td>
</tr>
</tbody>
</table>

Note. (A) (B) (C) (D) are types of disagreement between Algo+ and EPFL grader (illustrated in Table 4).

The two graders agree on scores for 2490 (79%) cases in the first exercise and 2671 (54%) in the second exercise but differ in the rest of cases (21% and 46% for Exercise 1 and Exercise 2, respectively). On the other hand, three types of differences were identified:

- **Case 1:** Submissions judged correct by EPFL grader but somewhat correct or erroneous by Algo+. This case represented 77 (2%) and 411 (8%) submissions in the first and the second exercise, respectively (disagreement type A and B illustrated in Table 4).

- **Case 2:** Submissions judged somewhat correct by EPFL grader but erroneous by Algo+. This case represented 5 (0.1%) and 1696 (34%) cases in exercise 1 and exercise 2, respectively (disagreement type C illustrated in Table 4).

- **Case 3:** Submissions judged erroneous by EPFL grader but somewhat correct by Algo+. This case represented 557 (18%) and 136 (2%) submissions in the exercise 1 and exercise 2, respectively (disagreement type D illustrated in Table 4).

To illustrate each case of disagreement between Algo+ and EPFL grader, we have chosen some illustrative samples from the two exercises in Table 4. According to these examples, we can highlight that EPFL grader awards high scores than Algo+ when the submitted program is not frequent because Algo+ does not recognize it among referent solutions. In the other case, when Algo+ awards high scores than EPFL grader, is due to the fact that Algo+ awards scores even if the submitted program does not give a correct output.
Table 4. Some examples that illustrate differences in scoring between Algo+ and EPFL grader

<table>
<thead>
<tr>
<th>Disagreement</th>
<th>Exercise1</th>
<th>Exercise2</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>int tmp = 0; if (chapeau) {</td>
<td>if (chapeau) {</td>
<td>Correct programs that were not frequent and thus Algo+ did not assigned the total score</td>
</tr>
<tr>
<td></td>
<td>tmp = c; if ((homme) &amp; &amp; (moustaches) &amp; &amp; (lunettes)) {</td>
<td>(&quot;le Professeur Violet&quot;); else {}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>c=b; if (moustaches) {</td>
<td>&quot;Il n'existe pas&quot;); else {}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>b=tmp; cout &lt;&lt; &quot;Il n'existe pas&quot;;} else {}</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>tmp = b; if (moustaches) {</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>b=a; if ((homme) &amp; &amp; (lunettes)) {</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>a=tmp; cout &lt;&lt; &quot;le Colonel Moutarde&quot;;} else {</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>if (lunettes) {</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>if (homme) {</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>cout &lt;&lt; &quot;le Reverend Olive&quot;;} else {</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>cout &lt;&lt; &quot;Mme Leblanc&quot;;} else {}</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>cout &lt;&lt; &quot;Mlle Rose »;}])</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(EPFL grader=30; Algo+=12.5)</td>
<td>(EPFL grader=50; Algo+=20)</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>swap(a,b); if (homme &amp; &amp; lunettes) {</td>
<td>if (homme &amp; &amp; lunettes) {</td>
<td>Correct programs that were not frequent and are not similar to any referent solution. Therefore Algo+ did not recognize this submission and assigned the lowest score (0)</td>
</tr>
<tr>
<td></td>
<td>swap(b,c); if (homme) {</td>
<td>(&quot;le Professeur Violet&quot;}; else {}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>swap(a,b); else {</td>
<td>(&quot;erreur&quot;); else {}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>swap(c,b); if (homme) {</td>
<td>(&quot;erreur&quot;); else {}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>cout &lt;&lt; &quot;le Colonel Moutarde&quot;;} else {</td>
<td>(&quot;erreur&quot;); else {}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>cout &lt;&lt; &quot;le Reverend Olive&quot;;} else {</td>
<td>(&quot;erreur&quot;); else {}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>cout &lt;&lt; &quot;Mme Leblanc&quot;;} else {}</td>
<td>(&quot;erreur&quot;); else {}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>cout &lt;&lt; &quot;Mlle Rose »;}])</td>
<td>(&quot;erreur&quot;); else {}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(EPFL grader=30; Algo+=0)</td>
<td>(EPFL grader=50; Algo+=0)</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>a=b+a; if (chapeau) {</td>
<td>if (homme) {</td>
<td>Somewhat correct programs that were not frequent, consequently, Algo+ did not recognize them</td>
</tr>
<tr>
<td></td>
<td>b=a+(-b); if(chapeau){</td>
<td>(&quot;le Colonel Moutarde&quot;}; else {}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>c=(b+a)/2; else{</td>
<td>(&quot;le Reverend Olive&quot;}; else {</td>
<td></td>
</tr>
<tr>
<td></td>
<td>if(chapeau){</td>
<td>(&quot;erreur&quot;); else {}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>if (homme) {</td>
<td>(&quot;erreur&quot;); else {}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>cout &lt;&lt; &quot;Mme Leblanc&quot;;} else {}</td>
<td>(&quot;erreur&quot;); else {}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>cout &lt;&lt; &quot;Mlle Rose »;}])</td>
<td>(&quot;erreur&quot;); else {}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(EPFL grader=10; Algo+=0)</td>
<td>(EPFL grader=40; Algo+=0)</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>int nTemp; if (chapeau) {</td>
<td>if (chapeau) {</td>
<td>Programs that did not produce correct output (EPFL grader=0) but they have a part of correctness in their code. Algo+ awards an average score according to how much the program is similar to a referent correct program</td>
</tr>
<tr>
<td></td>
<td>nTemp = c; if (homme) {</td>
<td>cout &lt;&lt; &quot;le Professeur Violet&quot;;} else {}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>b = a; if (homme) {</td>
<td>else if (moustaches) {</td>
<td></td>
</tr>
<tr>
<td></td>
<td>c = b; else if (lunettes) {</td>
<td>&quot;le Colonel Moutarde&quot;;} else {}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>a = nTemp; if (homme) {</td>
<td>(&quot;erreur&quot;); else {}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>cout &lt;&lt; &quot;Mlle Rose&quot;;} else {}</td>
<td>(&quot;erreur&quot;); else {}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>cout &lt;&lt; &quot;le Reverend Olive&quot;;} else {}</td>
<td>(&quot;erreur&quot;); else {}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>cout &lt;&lt; &quot;Mme Leblanc&quot;;} else {}</td>
<td>(&quot;erreur&quot;); else {}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>else {cout &lt;&lt; &quot;Mme Leblanc&quot;;}</td>
<td>(&quot;erreur&quot;); else {}</td>
<td></td>
</tr>
</tbody>
</table>
Analyze

The correlations between Algo+ and EPFL grader were calculated using Spearman’s rho and showed a high correlation between the two systems in the first exercise ($r_s = 0.92$, $p < .001$) and less high correlation in the second exercise ($r_s = 0.59$, $p < .001$) compared to the first exercise. Overall, there was a positive correlation between Algo+ and EPFL grader scores.

When the two distributions scores of the two exercises were compared, we found that the high agreement in the first exercise (ICC = 0.88, 95% CI 0.86-0.89, $p < .001$) is due to the simplicity of the exercise where variability of solutions is less important (72 distinct correct solutions produced by students where only 33 submissions were frequent) compared to the second exercise (ICC = 0.51, 95% CI 0.48-0.55, $p < .001$) where students came up with 482 correct ways to solve the exercise with 119 were frequent. The disagreement between EPFL grader and Algo+ is mainly explained by the fact that scores assigned by Algo+ is based solely on the frequent correct solutions. The rest of correct solutions that are not frequent, Algo+ will not recognize them and thus awards low scores.

However, when we examined the ranges of marks we found similarities and differences. Algo+ and EPFL grader assess correct programs with the same accuracy but differently when programs are not correct. Results showed that Algo+ and EPFL grader assess using different assessment metrics. Algo+ assessment is based on the existence of each part of the program in the referent solution whereas EPFL grader on the output correctness. Consequently Algo+ attributes high scores compared to EPFL grader especially in the case of erroneous programs. In this case Algo+ gives some points for approximate programs even if they do not run. On the other side, EPFL grader does not award any points when the program does not run or it does not produce the correct outputs.

In other words, EPFL grader does not assess bad programs, it assesses only correct programs. It looks for the correctness of outputs while students can produce interesting programs but may they forgot some details due to carelessness or silly errors that make the program not-running or produce the wrong outputs.

Also, we analyzed the approach adopted by Algo+ that uses the $n$ most frequent solutions to assess students’ submissions. We have found that Algo+ has yielded coverage of 96% and 75% of students’ submissions in the first and second exercise, respectively. We analyzed the rest of the not covered submissions and we have found only 2.3% and 7.5% of submissions in the first and second exercise, respectively, that represent correct and not frequent submissions, and thus Algo+ did not assess them correctly. These isolated cases come from some students that proposed functionally correct but somewhat weird solutions.

Types of feedback

One of the key challenges of program grading is designing feedback to the student to help them understand defects in their program (Wilcox, 2016). Both Algo+ and EPFL grader provide a feedback when they assess students’ submissions. By the nature of the assessment (static and dynamic program analysis) several differences were found to be significant.

Algo+ provides feedback with two parts. The first part of feedback is the annotations of the instructor about the conception of the whole solution. This feedback is annotated by the instructor when a submission has been added into the referent solutions database. If the referent solution is correct, a general explanation of how it is structured. Likewise, if the solution contains a misconception and semantic errors, also an explanation of this misconception is presented for student. The second part of feedback is generated from the difference between the submitted program and the most similar correct referent solution. This difference shows student how s/he has to carry out to obtain the correct solution. It shows students the missed instructions and which modifications have to be performed to get a valid solution. Using the feedback on the most popular erroneous programs tries to help students that do not understand why their programs do not provide the correct answers even if they are syntactically correct.
However, the feedback provided by EPFL grader is based on check style process and test cases. The generated feedback informs students about the test case achieved and not achieved by presenting the output of the submitted program and the expected output. Moreover, penalties on each undesirable practice such as unread variable are also generated by the feedback.

**Speed and quality of assessment of Algo+**

As Algo+ assessment is based on the most popular submissions and needs instructor intervention to annotate them, we have tried to explore the notion of reaching performance - we want to know how many submissions all frequent solutions could appear to let Algo+ independent from instructor and hence ensure its efficiency. We have chosen the number of submissions as an independent variable rather than the time spent between submissions because the number of submissions can change one in a while during a MOOC session.

In this study, Algo+ started assessment session with two programs that represent referent solutions in each exercise. During submitting programs (Exercise 1, $n = 3130$ and Exercise 2, $n = 4914$) to Algo+, instructor has manually assessed 33 and 119 programs in Exercise 1 and Exercise 2 respectively. These programs were appeared twice and more during the MOOC session and that have been assessed by the instructor.

To analyze the mechanism of Algo+ in reaching performance and hence to determine how many times instructor will be solicited in each exercise, we have passed to Algo+ the same chronological submissions of EPFL grader real session and we calculated the number of frequent solutions in each of 50 passed submissions.

After the examination of the relationship between the number of submissions and the number of frequent submissions that can appear, we have found that this relation is linear at the beginning but after some number of submissions, the number of frequent programs becomes rapidly constant and reaches an asymptotic level. This is a characteristic of hyperbolic models. What they have in common is a specified maximum value of the dependent variable, and a description of when the maximum value is attained in respect of the independent variable. There are several different types of hyperbolic functions. Among those models, the Michaelis-Menten model of substrate uptake. It consists of studies of the growth rates of microbes under controlled conditions commonly examine the relation between some measure of growth rate, and the substrate concentration. In the following section, we explained how the Michaelis-Menten model is used and adapted in our context.

**Toward modeling the number of frequent submissions in MOOCs based on adapted enzyme kinetics model**

Algo+ adopts an approach of giving feedback based on the $n$ most common submissions that instructor has to annotate. An automated feedback is giving by Algo+ when redundant submissions were detected. Redundant submissions are submissions that have the same syntactic structure. In this section, we attempted to conceive and to develop a model to accomplish the goal of estimating the number of frequent submissions (the $n$ most common solutions) for a given exercise.

To fit a model, the functional form of the growth trajectory of the frequent submissions being modeled must be chosen. To select the appropriate functional form for our model, we identified a small number of model types that aligned well with our data (i.e., they featured an interpretable upper asymptote to estimate the number of frequent submissions) and then tested their fit to the data empirically. Specifically, we compared the empirical fit of three different models with differing functional forms including S-shaped and J-shaped growth trajectories and identified the Michaelis-Menten model as the best fitting. The Michaelis-Menten model posits a J-shaped growth line and was first formulated in the field of biochemistry to estimate the rate of enzyme reactions based on the concentration of a substrate (English et al., 2006).

Notationally, the Michaelis-Menten growth trajectory of the number of frequent submissions according to the total number of submissions can be written as:

$$N = N_0 + \frac{N_{\text{max submissions}}}{K + \text{submissions}}$$

(1)

In Equation 1, the three parameters of the Michaelis-Menten function are: the initial value ($N_0$), which is the initial number of solution defined by the instructor when the MOOC session does not started yet (submissions = 0), the maximum number of submissions ($N_{\text{max submissions}}$), and the half-saturation constant ($K$).
0); the rate parameter ($K$), which represents the point in submissions when the number of frequent submissions is halfway between the initial value ($N_0$) and the asymptote; and the asymptote ($N_{max}$), which characterizes the maximum value of the frequent submissions as the number of submissions approaches infinity.

Figure 6 shows the growth of the number of referent submissions during each number of submissions and the fitted line using the adapted MM model.

![Figure 6. Plot of the number of frequent submissions (N) detected by Algo+ in the two exercises and the fitted line using the adapted Michaelis-Menten model. Submissions are chronologically ordered](image)

**Conclusions and future work**

In this work, Algo+’s behavior was examined when placed in the context of MOOC by comparing it to EPFL grader experience. Despite the difference of the adopted assessment approach by each tool, a significant correlation existed between the two tools. Scoring results have shown that the EPFL grader awards scores according to the functionality of the submitted program while Algo+ awards scores according to the existence of correct codes even if the program does not run.

Regarding the relation between Algo+ and EPFL grader that we can notice through this study, is that both can complement each other very well. We can use EPFL grader to assess the functionality of the submitted code and Algo+ to assess non-functional programs.

In the light of these study, we have also explored the utility and limitations of the supervised approach used by Algo+ where instructor has to assign feedback and score only on the $n$ most frequent submissions. This conducted us to propose a model about the relationship between the number of submissions and the number of frequent submissions that can appear in a given programming assignment during a MOOC session. The capitalization of the $n$ most frequent submissions used by Algo+ for assessing is a process meant to build up a capital from the huge number of submissions in a MOOC session, in order to develop the ability of learning programming by making (knowledge and competencies) available to other institutions and students.

For future work, we still need to assess how much Algo+ can improve students’ achievement in programming and the efficiency of the given feedback. This study has opened up new research areas for further improvements and future work. An automated grader that combined the two approaches can be developed. Also, it would be worthwhile to test the proposed model with different exercises to test the efficiency of predicting the critical number of submissions required to achieve the asymptotic level of frequent submissions and hence we could know, in an exercise, after how many submissions the $n$ most frequent submissions will appear.
Acknowledgments

We thank Ian Anthony Flitman and Jean-Cédric Chappelier from EPFL for their support and for providing us the EPFL grader data.

References


Learning analytics (LA) focuses on the analysis and interpretation of data related to learners’ profiles, learning contexts, and learners’ behaviours and interactions (Hwang, Chu, & Yin, 2017). One major objective of LA is to identify at-risk students and offer timely assistance (Selater, Peasgood, & Mullan, 2016). Various types of data on learning management systems (LMS) or virtual learning environments (VLE) and different classification methods have been investigated for predicting the potential failure or success of students. There are also large-scale successful implementations of LA systems, such as Course Signals (Arnold & Pistilli, 2012) and OU Analyse (Kuzilek, Hlosta, Herrmannova, Zdrahal, & Wolff, 2015). Identifying the struggling students at an early stage and providing remedial action to those potential failures as soon as possible have been recognised as being of paramount importance (Wolff, Zdrahal, Herrmannova, Kuzilek, & Hlosta, 2014; Alamuddin, Brown, & Kurzweil, 2016).

At present, most LA projects are implemented at the institutional or course level. For instructors, practising LA can be costly, as it involves handling a large number of variables and seeking inter-departmental collaboration, or even institutional support, to obtain the required LMS or VLE data. The reliance on big data at the institutional level was reported as a challenge for successful implementation of LA (Shewmaker & Fang, 2016). Despite the need for sufficient technical and human resources, the effort involved in data processing can be as high as 85% of the cost of LA implementation (Bienkowski, Feng, & Means, 2012). Gašević, Dawson, Rogers and Gašević (2016) stressed the problem of targeting at a one-size-fits-all LA solution, and the importance of taking into consideration the instructional conditions, which vary across disciplines, and instructors’ preferences, in LA implementation. They proposed course-specific models which provide better insights into the improvement of instructional practice, which implies that further effort needs to be devoted by instructors in addition to the institution.

At the course level, clickers (or classroom response systems) have long been adopted to promote interaction in classrooms. They are typically used to collect attendance, summative assessment, or survey data from students in classes (Fies & Marshall, 2006). Instructors can then obtain students’ learning progress in real time and react to it. Despite a few instances (e.g., Chien, Lee, Li, & Chang, 2015; Majumdar & Iyer, 2016), clickers have not been a common source of data for LA. This could be due to the high cost of their use, which traditionally requires a response device for each student and a receiver (Blasco-Arcas, Buil, Hernández-Ortega, & Sese, 2013). With the advance of mobile technology, the function of clickers can be performed using students’ own mobile devices and free web-based services, so that relevant data can be obtained at a much lower cost (Glassman, 2015).
Following the identification of students at risk, the implementation of intervention has been claimed to be the largest challenge for LA research and practice (Rienties, Cross, & Zdrahal, 2017). Campbell and Oblinger (2007) proposed a five-step analytics model — capture, report, predict, act and refine — for summarising the various stages of LA. Most research efforts have been focused on building an accurate prediction model (the first three steps). While it has also been emphasised in recent years that meaningful LA data should be returned to learners — closing the loop of the LA cycle (e.g., Clow, 2012; Corrin et al., 2016) — the intervention strategies and impact (the last two steps) have rarely been discussed in detail (Rienties et al., 2016; Rienties et al., 2017). This could be partially attributed to the limited expertise and experience of some instructors in interpreting LA outputs and taking respective actions (Corrin et al., 2016; Dazo, Stepanek, Chauhan, & Dorn, 2017). There is a need to help instructors to be involved in LA practises and perform intervention.

This paper reports a study on employing LA in a business quantitative methods course in order to identify at-risk students and conduct intervention. It illustrates the building of a multi-stage prediction model for students at risk, featuring the use of in-class clicker data together with students’ prior learning data. In addition, the performance of a linear regression model and a logistic regression model for classification is compared to show their accuracy in predicting students’ study results. Based on the prediction analysed from the collected data, interventions were made to the students who were potential failures through various strategies. The paper also presents how intervention was practised at a low cost to minimise the instructors’ effort, as well as the effect and impact of the intervention.

**Literature review**

**Identification of at-risk students**

Prediction of student performance is one of the major research objectives in both LA and educational data-mining communities (Papamitsiou & Economides, 2014), through which students at risk of failing a course can be identified. For example, Osborn (2001) made use of a survey, with variables such as learning motivation and need for help, to determine students who were at risk of abandoning a course. Allen, Higgs and Holloway (1988) identified early the strong relevance of students’ pre-admission GPA and pre-requisite GPA to their need for academic support. Alston, Lane and Wright (2014) also found that students’ exam results for courses taken early in a programme correlate with their success in graduating.

LMS or VLE has been a main source of data for tracking students’ engagement through their activeness in the system. Morris, Finnegan and Wu (2005) determined that the extent of students’ participation in online learning activities predicts their study achievement. Macfadyen and Dawson (2010) identified 15 variables which had a significant correlation with student final grades, such as the numbers of discussion posts, emails and completed assignments. Also, Wolff, Zdrahal, Nikolov and Pantucek (2013) further revealed that the clicks in VLE, as a type of coarse-grained data, can reliably predict student failure.

Despite the promise of these variables as effective predictors in relevant studies, there are limitations in their use. As illustrated in Marbouti, Diefes-Dux and Madhavan (2016), the effectiveness of the variables can vary a great deal for courses with diverse learning objectives, activities and assessments. Also, those variables which rely heavily on data on LMS/VLE may not be suitable for face-to-face courses with learning activities happening mainly in classrooms.

**Use of clickers**

Clickers have long been used to promote interaction in classrooms (Dufresne, Wenk, Mestre, Gerace, & Leonard, 1996; Mestre, Gerace, Dufresne, & Leonard, 1997), and have shown a broad range of benefits. As reviewed in Fies and Marshall (2006), the major benefits include “greater student engagement, increased student understanding of complex subject matter, heightened discussion and interactivity, increased student awareness of individual levels of comprehension, and increased teacher insight into student difficulties” (p. 103). In the past, the use of clickers was relatively costly. It consisted of transmitters that students used to send responses, receivers that collected inputs, and computer software that interpreted and aggregated these responses in real time (Fies & Marshall, 2006). With the popularity of mobile devices and free cloud-based tools, the functions of a clicker can be provided on students’ mobile devices, free mobile apps and the support of cloud-based analytical tools (Glassman, 2015).
There has been limited work on applying LA to the formative assessment data collected by clickers or computer-assisted tests. Two relevant studies are Porter, Zingaro and Lister (2014) and Tempelaar, Rienties and Giesbers (2015). The former predicted student success in a computer science course using in-class clicker data and found that students’ exam performance was correlated with their in-class clicker scores; and at-risk students could be identified by evaluating students’ progress every three weeks. The latter involved an empirical research study for a quantitative methods course and found that formative assessment is the most predictive indicator of students’ success.

**Intervention strategies**

Intervention is regarded as the final stage of the LA cycle — to take remedial actions on learners (Clow, 2012). It is also claimed as the largest challenge ahead of LA research and practice (Rienties et al., 2017) in terms of uncovering the effect and impact of different intervention strategies on at-risk students identified by using LA. In some LA practices, instances of intervention were suggested. For Course Signals, Arnold and Pistilli (2012) only listed examples, such as showing an alert indicator to students, email reminders, text messages and face-to-face meetings with instructors. Lonn, Aguilar and Teasley (2015) briefly illustrated the meetings with students by academic advisors as an intervention in the context of a summer bridge programme. Also, Rienties et al. (2016) stated that there is a lack of systematic organisation of intervention approaches for the reference of instructors and administrators.

Intervention also presents challenges for students and instructors. Wise (2014) explained that LA visualisation and signals for students require them to have strong metacognitive skills to utilise the analytics for reflection and self-regulation. Students at risk of failure may be relatively weak in those skills and thus have difficulty in engaging with intervention. For instructors, taking remedial actions based on LA data implies a time cost, which is more important in larger courses where the time and resources of instructors may be too prohibitive for them to take proactive intervention (Corrin et al., 2016).

**Methodology**

This study aims to examine the effectiveness of in-class clicker data for early prediction of students at risk of failing a course. The clicker data were collected, together with other data which have been shown to be effective in previous studies, to build a prediction model. This research also proposes a number of intervention methods and analyses their strengths and limitations, based on the experience of LA practice for the course in this study. The proactive use of intervention methods in relation to the risk levels of students and the stages of a course is also suggested. Specifically, this study focuses on the following research questions:

- How effective is clicker data, compared with other data, for identifying students at risk of failing examinations?
- What is the difference in performance between linear regression and logistic regression for identifying the at-risk students?
- How effective is the proactive intervention strategy? Do intervention frequency and peer groups affect the pass rate of at-risk students?

**Data**

This study was conducted on a first-year undergraduate face-to-face course (Quantitative Methods for Decision Making) at the Open University of Hong Kong. The course is concerned with the application of statistical and mathematical modelling in business, and requires students to learn and apply quantitative models for driving business decisions. It is a compulsory course for the Bachelor of Business Administration. Students on the course had been found to be more likely to fail in the exam because of having insufficient background knowledge. There were 564 and 511 students enrolled in the course in 2016 and 2017 respectively. Delivered through 12 mass lectures and small-group tutorial classes over 12 weeks, the course covers six topics: ANOVA, linear and multiple regression, time series, decision analysis, project management, and linear programming. The first two topics are closely related to its pre-requisite course (Business Statistics), while the remaining topics are relatively independent and only loosely associated with the former ones.
Most of the existing LA research on finding at-risk students has analysed LMS or VLE tracking data. Gašević et al. (2016) advised that, due to the diverse nature of courses, the instructional condition should be carefully considered before applying LA to LMS data — a practice which demands additional effort from the instructors. In addition, LMS data need to be collected regularly (e.g., once a week) from the LMS administrator and substantial energy must then be spent in cleaning and processing the data. Since the course in question is a face-to-face course delivered through lectures and tutorials, the usage of LMS has been low. Therefore, we decided not to include LMS data in the prediction model. Instead, we utilised a self-designed classroom response system to collect student in-class feedback, which provides more valuable data for predictive analytics.

According to previous research (e.g., Tempelaar et al., 2015), two types of data have been reported that have the most dominant impact on the prediction of students who are potential failures — namely, students’ demographic data (including the pre-requisite exam score) and the formative/summative assessments, which are considered the most predictive indicators. Both types of data can be obtained with minimal effort — the demographic information is usually available for instructors or otherwise can be collected through student questionnaires, while the formative and summative assessment results can be collected as the course proceeds.

We employed Google Forms and Google Sheets to develop a low-cost classroom response system. They are free of charge and work on any devices with internet connection. All students on the course had their own mobile phones or tablets for using Google Forms in the classroom, and had high digital literacy in general without the need for extra training in using the classroom response system. Towards the end of each tutorial, students were asked to demonstrate their understanding of the concepts by answering three to five multiple-choice questions, including conceptual and practical questions. Students could use their own mobile devices to connect to Google Forms and answer the questions. The statistics on the answers collected were shown in real time and the instructor subsequently explained the correct answers to the students in class. Google Sheets was used for data analysis, in particular for linear regression and logistic regression, for the prediction of students at-risk.

As shown in Table 1, the students’ demographics and academic information (such as gender, academic standing and the exam score in the pre-requisite course) were collected. In addition, students were asked to fill in a questionnaire during the first tutorial to specify the grades they aimed at and their interest in mathematics, to show their extrinsic and intrinsic learning motivations respectively. The course data in the 2016 cohort (N = 160) were then used as a training set to predict the at-risk students in the 2017 cohort (N = 83).

| Table 1. List of attributes of training/testing data for at-risk prediction |
|-----------------|---------------------------------|
| **Attribute name** | **Description** |
| EXAM            | Examination score in the course |
| RETAKE          | Flag to indicate retaking of the course (“Yes” or “No”) |
| PRE_EXAM*       | Examination score in the pre-requisite course |
| GENDER          | Student’s gender (“M” or “F”) |
| INTEREST*       | Degree of interest in the course (“None,” “Little,” “Fair,” and “Great”) |
| EXPECT*         | Student’s expected course grade (mapped to the minimum score of the corresponding grade) |
| ASSIGNMENT      | Assignment score |
| QUIZ            | In-class quiz score |
| ATTEND          | Cumulative attendance rate (by week) |
| CLICKER         | Cumulative in-class test scores (by week) |

*Note.* Attributes with missing data.

The missing values were imputed by multiple imputation based on Gibbs sampling, where missing values on continuous variables (PRE_EXAM and EXPECT) and categorical variables (INTEREST) with more than two levels were replaced by predictive mean matching and polytomous logistic regression respectively. The imputation model includes the attributes with missing values and two related attributes: GENDER and RETAKE. Thirty imputation datasets were generated for separate analysis and the statistical results were then combined by averaging.

**Procedure**

To select the attributes for building the prediction model for identifying at-risk students, correlation analysis and the Kruskal-Wallis test were conducted to evaluate the usefulness of the attributes for predicting the exam score. Table 2 shows the correlations between continuous variables. ATTEND and CLICKER have cumulative figures...
for the 12 study weeks. The table also shows the correlations for the last week and comparable results were obtained in other study weeks. The correlations of EXAM with all other continuous variables had a moderately positive relationship and were statistically significant, except for EXPECT that was relatively weak and positive. This suggests that all the continuous variables, except EXPECT, would be good predictor variables for the exam score. The pairwise correlations between predictive variables also indicated a weak to moderate positive relationship, except between ATTEND and CLICKER. A strong correlation (0.9227) between ATTEND and CLICKER indicates that they are highly correlated predictors and including both in the prediction model will result in a multicollinearity problem. Given a strong correlation with CLICKER, ATTEND was not used to build the prediction model. Table 3 gives the Kruskal-Wallis test results between EXAM and GENDER, RETAKE, and INTEREST. Among the three tests, the test result of EXAM and GENDER is statistically insignificant and therefore GENDER was not included in the prediction model.

<table>
<thead>
<tr>
<th>EXAM</th>
<th>PRE_EXAM</th>
<th>EXPECT</th>
<th>ASSIGNMENT</th>
<th>QUIZ</th>
<th>ATTEND</th>
<th>CLICKER</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>0.6724</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EXPECT</td>
<td>0.2936</td>
<td>0.4532</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASSIGNMENT</td>
<td>0.6618</td>
<td>0.4663</td>
<td>0.0947*</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QUIZ</td>
<td>0.6749</td>
<td>0.5386</td>
<td>0.2266</td>
<td>0.6102</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>ATTEND</td>
<td>0.4963</td>
<td>0.2621</td>
<td>0.0541*</td>
<td>0.4575</td>
<td>0.4263</td>
<td>1</td>
</tr>
<tr>
<td>CLICKER</td>
<td>0.5488</td>
<td>0.3465</td>
<td>0.2097**</td>
<td>0.4816</td>
<td>0.4970</td>
<td>0.9227</td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
<td>(0.0000)</td>
</tr>
</tbody>
</table>

Note. *p > .10; **p > .05; two-tailed test.

<table>
<thead>
<tr>
<th>EXAM</th>
<th>GENDER</th>
<th>RETAKE</th>
<th>INTEREST</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.4684</td>
<td>2.7678*</td>
<td>18.15***</td>
<td></td>
</tr>
<tr>
<td>(0.2256)</td>
<td>(0.09618)</td>
<td>(0.0004)</td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .10; ***p < .01.

After the key attributes were identified, we constructed a five-stage prediction model by incorporating the ongoing formative and summative assessment data as follows.

Stage 1 (1st week): RETAKE, PRE_EXAM, INTEREST, EXPECT
Stage 2 (3rd week): +ATTEND + CLICKER
Stage 3 (6th week): +ASSIGNMENT
Stage 4 (9th week): +QUIZ
Stage 5 (12th week): Update for CLICKER

Several popular classification methods were studied by Jayaprakash, Moody, Lauria, Regan and Baron (2014) for identifying at-risk students. It was reported that the recall and precision rate of the various classifiers being studied were empirically comparable. For LA researchers, logistic regression is a common choice for classification tools. Based on the finding of Jayaprakash et al. (2014), we chose to adopt logistic regression as the baseline of our classification methods. Besides its simplicity and popularity, logistic regression can reduce the chance of over-fitting and provide a probability of failure for each case so that the students can be ranked according to the risk level. However, we also employed the linear regression with prediction intervals for the purpose of comparison.

Hierarchical logistic regression (LG) and hierarchical linear regression (LR) models were applied to develop prediction models for predicting the at-risk level of students. Both hierarchical regression models consisted of five stages for predicting the students’ at-risk level, as below.
Stage 1 (Week 0) \[ AT\_RISK \text{ or } EXAM = 12_0 + 12_1 \text{RETAKE} + 12_2 \text{PRE_EXAM} + 12_3 \text{INTEREST} + 12_4 \text{EXPECT} + e \]

where \( AT\_RISK = \begin{cases} \text{YES} & \text{if } EXAM < \text{Passing score} \\ \text{NO} & \text{if } EXAM \geq \text{Passing score} \end{cases} \)

Stage 2 (Week 3) 

\[ AT\_RISK \text{ or } EXAM = 12_0 + 12_1 \text{RETAKE} + 12_2 \text{PRE_EXAM} + 12_3 \text{INTEREST} + 12_4 \text{EXPECT} + 12_5 \text{CLICKER}_1 + e \]

where \( \text{CLICKER}_1 \) is the cumulative clicker score at week 3.

Stage 3 (Week 6) 

\[ AT\_RISK \text{ or } EXAM = 12_0 + 12_1 \text{RETAKE} + 12_2 \text{PRE_EXAM} + 12_3 \text{INTEREST} + 12_4 \text{EXPECT} + 12_5 \text{CLICKER}_2 + 12_6 \text{ASSIGNMENT} + e \]

where \( \text{CLICKER}_2 \) is the cumulative clicker score at week 6.

Stage 4 (Week 9) 

\[ AT\_RISK \text{ or } EXAM = 12_0 + 12_1 \text{RETAKE} + 12_2 \text{PRE_EXAM} + 12_3 \text{INTEREST} + 12_4 \text{EXPECT} + 12_5 \text{CLICKER}_3 + 12_6 \text{ASSIGNMENT} + 12_7 \text{QUIZ} + e \]

where \( \text{CLICKER}_3 \) is the cumulative clicker score at week 9.

Stage 5 (Week 12) 

\[ AT\_RISK \text{ or } EXAM = 12_0 + 12_1 \text{RETAKE} + 12_2 \text{PRE_EXAM} + 12_3 \text{INTEREST} + 12_4 \text{EXPECT} + 12_5 \text{CLICKER}_4 + 12_6 \text{ASSIGNMENT} + 12_7 \text{QUIZ} + e \]

where \( \text{CLICKER}_4 \) is the cumulative clicker score at week 12.

The LG model can predict the at-risk probability of students (i.e., the probability of failing the exam), while the LR model can predict their level of exam score. In order to develop an effective intervention strategy, the students were divided into LOW and HIGH at-risk levels for selecting appropriate intervention methods. The at-risk levels can be determined by setting probability thresholds for the predicted probability of the LG model and the level of significance of the prediction intervals in the LR model. A probability threshold pair \((\alpha_1, \alpha_2)\) where \(\alpha_1 < \alpha_2\) was defined for classifying students as LOW and HIGH at-risk levels based on the models’ predictions as follows:

**LG model:**

\[ AT\_RISK \text{ LEVEL } = \begin{cases} \text{LOW} & \text{if } \alpha_1 \leq \text{Prob}(AT\_RISK) < \alpha_2 \\ \text{HIGH} & \text{if } \alpha_2 \leq \text{Prob}(AT\_RISK) \end{cases} \]

where \( \text{Prob}(AT\_RISK) \) is the predicted at-risk probability.

**LR model:**

\[ AT\_RISK \text{ LEVEL } = \begin{cases} \text{LOW} & \text{if } Pl_\alpha(\alpha_2) \leq \text{Passing score} < Pl_\alpha(\alpha_1) \\ \text{HIGH} & \text{if } Pl_\alpha(\alpha_1) < \text{Passing score} \end{cases} \]

where \( Pl_\alpha(\alpha) \) is the lower limit of the \( \alpha \) prediction interval.

Tables 4 and 5 summarise the LG and LR models generated from the training dataset. Dichotomous variables (RETAKE) and ordinal variables (INTEREST) are transformed to dummy coding and polynomial contrasts: Linear (L), Quadratic (Q), and Cubic (C) respectively. The coefficient of all attributes except INTEREST and RETAKE are statistically significant. The predictive power of both the hierarchical regression models can be substantially improved by adding attributes in stages, except the last two stages. This is because regression models in Stages 4 and 5 have the same set of predictors and differ by the week of the clicker score being used in the prediction models. Despite moderate correlations between the predictive variables, the variance inflation factors (VIF) of predictors do not indicate multicollinearity issues in the hierarchical regression models. A tenfold cross-validation with 20 repetitions was also conducted to test the LG and LR models. The out of sample prediction errors of the models were measured by the mean squared error (MSE) on the training and testing sets and the cross-validation percentage errors of the models are included in the tables. The cross-validation percentage errors for the LG and LR models are small, ranging from 0.89% to 2.20% and 3.22% to 3.69% respectively. With the small percentage errors, both models should be valid for at-risk prediction.
### Table 4. LG model (standard errors in parentheses)

<table>
<thead>
<tr>
<th>Regression</th>
<th>Stage 1</th>
<th>Stage 2</th>
<th>Stage 3</th>
<th>Stage 4</th>
<th>Stage 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.9067</td>
<td>2.9466</td>
<td>7.8908***</td>
<td>10.6529***</td>
<td>10.7802***</td>
</tr>
<tr>
<td>RETAKE (YES)</td>
<td>1.6767</td>
<td>0.7394</td>
<td>-1.6639</td>
<td>-2.9463</td>
<td>-2.9208</td>
</tr>
<tr>
<td>PRE_EXAM</td>
<td>-0.0657***</td>
<td>-0.0686***</td>
<td>-0.0507***</td>
<td>-0.0454***</td>
<td>-0.0431***</td>
</tr>
<tr>
<td>INTEREST (L)</td>
<td>-0.4484</td>
<td>-0.6614</td>
<td>-1.0504</td>
<td>-1.0980</td>
<td>-0.9403</td>
</tr>
<tr>
<td>INTEREST (Q)</td>
<td>0.2664</td>
<td>0.3158</td>
<td>0.6521</td>
<td>0.1781</td>
<td>0.0800</td>
</tr>
<tr>
<td>INTEREST (C)</td>
<td>0.0507</td>
<td>-0.3673</td>
<td>-0.6734</td>
<td>-0.3488</td>
<td>-0.2424</td>
</tr>
<tr>
<td>EXPECT</td>
<td>-0.0065</td>
<td>-0.0033</td>
<td>-0.0224</td>
<td>-0.0376</td>
<td>-0.0420</td>
</tr>
<tr>
<td>CLICKER</td>
<td>-3.5337***</td>
<td>-5.6859***</td>
<td>-4.458*</td>
<td>-4.226*</td>
<td></td>
</tr>
<tr>
<td>ASSIGNMENT</td>
<td>-0.0586***</td>
<td>-0.0454***</td>
<td>-0.0455***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QUIZ</td>
<td>1.3215</td>
<td>0.0402</td>
<td>0.0424</td>
<td>0.0417</td>
<td></td>
</tr>
</tbody>
</table>

Akaike Information Criterion (AIC) 108.53 102.23 90.598 85.161 86.287
Cross-validation Error (by MSE) 1.03% 0.89% 2.20% 2.14% 2.12%

Note. **p < .05; ***p < .01; two-tailed test.

### Table 5. LR Model (standard errors in parentheses)

<table>
<thead>
<tr>
<th>Regression</th>
<th>Stage 1</th>
<th>Stage 2</th>
<th>Stage 3</th>
<th>Stage 4</th>
<th>Stage 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>RETAKE (YES)</td>
<td>-12.3885</td>
<td>-5.9796</td>
<td>8.2252</td>
<td>9.1734</td>
<td>9.1631</td>
</tr>
<tr>
<td>PRE_EXAM</td>
<td>0.7359***</td>
<td>0.6993***</td>
<td>0.5041***</td>
<td>0.4121***</td>
<td>0.4072***</td>
</tr>
<tr>
<td>INTEREST (L)</td>
<td>2.9076</td>
<td>3.0232</td>
<td>4.8326</td>
<td>5.5611*</td>
<td>4.8628</td>
</tr>
<tr>
<td>INTEREST (Q)</td>
<td>4.6771</td>
<td>4.39641</td>
<td>3.5961</td>
<td>3.7895</td>
<td>3.8111</td>
</tr>
<tr>
<td>INTEREST (C)</td>
<td>-2.4291</td>
<td>-4.2704</td>
<td>-4.3749</td>
<td>-4.5003</td>
<td>-3.9873</td>
</tr>
<tr>
<td>EXPECT</td>
<td>-0.9355</td>
<td>-0.0120</td>
<td>0.8163</td>
<td>0.8692</td>
<td>0.5019</td>
</tr>
<tr>
<td>CLICKER</td>
<td>19.4314***</td>
<td>23.2053***</td>
<td>20.9657***</td>
<td>19.5714***</td>
<td></td>
</tr>
<tr>
<td>ASSIGNMENT</td>
<td>4.0521***</td>
<td>0.3669***</td>
<td>0.3668***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QUIZ</td>
<td>0.2294**</td>
<td>0.233***</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R² 0.5088 0.5693 0.6649 0.6875 0.6819
F-statistic 26.42*** 28.7*** 37.45*** 36.6*** 35.73***
Cross-validation Error (by MSE) 3.69% 3.22% 3.41% 3.69% 3.41%

Note. *p < .10; **p < .05; ***p < .01; two-tailed test.
Analysis of model performance

The testing dataset for the 2017 cohort (N = 83) was used to evaluate the performance of the at-risk models at six probability threshold pairs from (0.25, 0.5) to (0.5, 0.75) increased by increments of 0.5. The performance of the prediction models was evaluated using the recall, precision, accuracy and false positive (FP) rates derived from the true positive (TP), true negative (TN), false positive (FP), and false negatives (FN) of the confusion matrix:

Recall = \( \frac{TP}{TP + FN} \); Precision = \( \frac{TP}{TP + FP} \); Accuracy = \( \frac{TP + TN}{TP + TN + FP + FN} \); FP rate = \( \frac{FP}{TN + FN} \)

Recall represents the ability of the prediction models to detect at-risk students; precision represents the correctness of the at-risk detection result; accuracy represents the overall accuracy of the prediction models in classifying at-risk and non-at-risk students; and the FP rate represents the false alarms raised by the prediction models.

The prediction models classify the risk levels of students from a low to high risk. Tables 6 and 7 summarise the overall performance of the prediction models in identifying at-risk students either at low or high levels. Recall, precision, and accuracy are gradually increased with the stage, while the FP rate is gradually decreased. For the LG model, increasing the probability thresholds significantly reduced the recall but raised slightly the precision of the prediction model. In contrast, increasing the probability thresholds for the LR model significantly increased the recall but only reduced the precision slightly. The accuracy of the prediction models was comparable, but the FP rate was higher in the LR model. Improvement in the recall also resulted in the increase of false alarms in the LR model. Among the five stages, the performance of the prediction models in Stages 4 and 5 were quite close. Given that the predictions of the last two stages were made from the same set of predictive variables with clicker score updated and only two additional clicker tests conducted, the performances of the prediction models were only slightly different in these stages.

### Table 6. Performance of the LG model

<table>
<thead>
<tr>
<th>Probability threshold</th>
<th>0.25/0.5 (%)</th>
<th>0.3/0.55 (%)</th>
<th>0.35/0.6 (%)</th>
<th>0.4/0.65 (%)</th>
<th>0.45/0.7 (%)</th>
<th>0.5/0.75 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recall</td>
<td>43.7</td>
<td>39.8</td>
<td>34.9</td>
<td>31.4</td>
<td>27.3</td>
<td>22.4</td>
</tr>
<tr>
<td>Precision</td>
<td>45.3</td>
<td>47.3</td>
<td>48.0</td>
<td>48.7</td>
<td>48.4</td>
<td>44.8</td>
</tr>
<tr>
<td>Accuracy</td>
<td>72.1</td>
<td>73.1</td>
<td>73.3</td>
<td>73.5</td>
<td>73.7</td>
<td>73.1</td>
</tr>
<tr>
<td>FP Rate</td>
<td>18.3</td>
<td>15.6</td>
<td>13.8</td>
<td>12.2</td>
<td>10.6</td>
<td>9.8</td>
</tr>
<tr>
<td>Stage 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recall</td>
<td>48.6</td>
<td>42.4</td>
<td>37.5</td>
<td>33.8</td>
<td>29.8</td>
<td>25.4</td>
</tr>
<tr>
<td>Precision</td>
<td>46.9</td>
<td>46.4</td>
<td>46.5</td>
<td>48.1</td>
<td>47.2</td>
<td>47.6</td>
</tr>
<tr>
<td>Accuracy</td>
<td>72.7</td>
<td>72.6</td>
<td>72.9</td>
<td>73.6</td>
<td>73.4</td>
<td>73.7</td>
</tr>
<tr>
<td>FP Rate</td>
<td>19.1</td>
<td>17.2</td>
<td>17.5</td>
<td>12.9</td>
<td>11.8</td>
<td>9.9</td>
</tr>
<tr>
<td>Stage 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recall</td>
<td>58.3</td>
<td>54.4</td>
<td>48.6</td>
<td>45.6</td>
<td>42.9</td>
<td>39.4</td>
</tr>
<tr>
<td>Precision</td>
<td>59.7</td>
<td>59.9</td>
<td>58.1</td>
<td>58.5</td>
<td>59.4</td>
<td>59.8</td>
</tr>
<tr>
<td>Accuracy</td>
<td>79.5</td>
<td>79.2</td>
<td>78.1</td>
<td>78.0</td>
<td>78.2</td>
<td>78.0</td>
</tr>
<tr>
<td>FP Rate</td>
<td>13.3</td>
<td>12.4</td>
<td>11.9</td>
<td>11.0</td>
<td>9.9</td>
<td>9.0</td>
</tr>
<tr>
<td>Stage 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recall</td>
<td>62.9</td>
<td>60.3</td>
<td>58.3</td>
<td>57.5</td>
<td>55.4</td>
<td>53.7</td>
</tr>
<tr>
<td>Precision</td>
<td>66.1</td>
<td>65.8</td>
<td>66.8</td>
<td>68.5</td>
<td>70.8</td>
<td>72.1</td>
</tr>
<tr>
<td>Accuracy</td>
<td>82.4</td>
<td>82.0</td>
<td>82.1</td>
<td>82.5</td>
<td>82.9</td>
<td>83.0</td>
</tr>
<tr>
<td>FP Rate</td>
<td>11.0</td>
<td>10.6</td>
<td>9.8</td>
<td>9.0</td>
<td>7.8</td>
<td>7.1</td>
</tr>
<tr>
<td>Stage 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recall</td>
<td>63.2</td>
<td>61.3</td>
<td>59.0</td>
<td>57.1</td>
<td>54.9</td>
<td>53.7</td>
</tr>
<tr>
<td>Precision</td>
<td>66.1</td>
<td>66.7</td>
<td>67.3</td>
<td>67.1</td>
<td>68.7</td>
<td>71.1</td>
</tr>
<tr>
<td>Accuracy</td>
<td>82.4</td>
<td>82.4</td>
<td>82.4</td>
<td>82.0</td>
<td>82.2</td>
<td>82.7</td>
</tr>
<tr>
<td>FP Rate</td>
<td>11.0</td>
<td>10.4</td>
<td>9.7</td>
<td>9.5</td>
<td>8.5</td>
<td>7.5</td>
</tr>
</tbody>
</table>

The recall and precision of the LG and LR models for probability threshold pairs (0.25, 0.5) and (0.5, 0.75) are compared in Figures 1 and 2. The figures suggest that the LR model outperforms the LG model in recall at all stages, but the LG model provides higher precision in the last two stages (Stages 4 and 5). As usual, the selection of probability threshold pairs for prediction models would be a trade-off between recall and precision. To avoid missing out at-risk students from intervention, FN would be less harmful than FP and, therefore, recall would be a more important criterion for selecting an at-risk prediction model than would precision. From this perspective, the LR model would be preferable to the LG model.
<table>
<thead>
<tr>
<th>Table 7. Performance of the LR model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability threshold</td>
</tr>
<tr>
<td>Stage 1 Recall</td>
</tr>
<tr>
<td>Precision</td>
</tr>
<tr>
<td>Accuracy</td>
</tr>
<tr>
<td>FP Rate</td>
</tr>
<tr>
<td>Stage 2 Recall</td>
</tr>
<tr>
<td>Precision</td>
</tr>
<tr>
<td>Accuracy</td>
</tr>
<tr>
<td>FP Rate</td>
</tr>
<tr>
<td>Stage 3 Recall</td>
</tr>
<tr>
<td>Precision</td>
</tr>
<tr>
<td>Accuracy</td>
</tr>
<tr>
<td>FP Rate</td>
</tr>
<tr>
<td>Stage 4 Recall</td>
</tr>
<tr>
<td>Precision</td>
</tr>
<tr>
<td>Accuracy</td>
</tr>
<tr>
<td>FP Rate</td>
</tr>
<tr>
<td>Stage 5 Recall</td>
</tr>
<tr>
<td>Precision</td>
</tr>
<tr>
<td>Accuracy</td>
</tr>
<tr>
<td>FP Rate</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Table 8. Recall and precision of low and high at-risk levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability threshold</td>
</tr>
<tr>
<td>Risk level*</td>
</tr>
<tr>
<td>Stage 1 Recall</td>
</tr>
<tr>
<td>Precision</td>
</tr>
<tr>
<td>Stage 2 Recall</td>
</tr>
<tr>
<td>Precision</td>
</tr>
<tr>
<td>Stage 3 Recall</td>
</tr>
<tr>
<td>Precision</td>
</tr>
<tr>
<td>Stage 4 Recall</td>
</tr>
<tr>
<td>Precision</td>
</tr>
<tr>
<td>Stage 5 Recall</td>
</tr>
<tr>
<td>Precision</td>
</tr>
</tbody>
</table>

Figure 1. Comparison of the recall at risk prediction models
To further explore the performance of the prediction models in classifying the at-risk levels of students, the recall and precision of the LR model were broken down by low and high at-risk levels (see Table 8). The results for probability threshold pairs (0.25, 0.5) and (0.5, 0.75) are shown in Figures 3 and 4. From the figures, the recall of high at-risk level increased with the stage, while the low at-risk level decreased when the stage increased. The precision in general also increased with the stage for both low and high at-risk levels. Overall, the performance of the prediction model would be improved by incorporating more predictive variables for measuring the students’ learning progress, and the prediction model effectively identifies more at-risk students as at the high at-risk level than the low at-risk level at the later stages.

**Figure 2.** Comparison of the precision of at-risk prediction models

**Figure 3.** Comparison of the recall of low/high at-risk levels (Hierarchical Linear Regression Model)
The proposed intervention strategy adopts a systematic approach to proactive advising, also known as “intrusive
advising” (Earl, 1988; Varney, 2007; Varney, 2012), which is described as one of the most effective advising
techniques for supporting at-risk students (Schwebel, Walburn, Jacobsen, Jerrolds, & Klyce, 2008). Proactive
advising advocates that advisors should establish a good rapport with students by taking the initiative to contact
them in a friendly manner. Proactive advising also emphasises taking preventive measure to connect students
before problems occur.

### Intervention strategies

Table 9. Pros and cons for the commonly-used intervention methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Email</td>
<td>• Least expensive</td>
<td>• Students may easily overlook the message due to too many spam emails</td>
</tr>
<tr>
<td></td>
<td>• Allows personalisation via mail merge</td>
<td></td>
</tr>
<tr>
<td>Phone call</td>
<td>• Good for emergency matters – two-way synchronous communications</td>
<td>• Students may not be available and sometimes feel offended</td>
</tr>
<tr>
<td>Instant messaging</td>
<td>• Preferred communication channel for many students</td>
<td>• More costly than email as it requires one-to-one communications</td>
</tr>
<tr>
<td>LMS post &amp; news</td>
<td>• Facilitates many-to-many asynchronous communications</td>
<td>• Requires students to login to the LMS and may overlook the posts and news</td>
</tr>
<tr>
<td>Group consultation</td>
<td>• Effective communication</td>
<td>• Usually needs making appointments in advance and expensive for instructors</td>
</tr>
<tr>
<td></td>
<td>• Good for timid students</td>
<td></td>
</tr>
<tr>
<td>Face-to-face consultation</td>
<td>• Effective communication</td>
<td>• Most expensive and usually needs to make appointments in advance</td>
</tr>
<tr>
<td></td>
<td>• One-to-one consultation</td>
<td></td>
</tr>
<tr>
<td>Video recording</td>
<td>• Effective instruction</td>
<td>• Substantial initial effort to record the instructions</td>
</tr>
<tr>
<td></td>
<td>• Not restricted by time</td>
<td></td>
</tr>
<tr>
<td>Peer review</td>
<td>• Encourages critical evaluation</td>
<td>• Requires good question design</td>
</tr>
<tr>
<td></td>
<td>• Students can learn from each other</td>
<td>• Often conducted in class</td>
</tr>
<tr>
<td>E-tutorial</td>
<td>• Supplementary instructions available 24/7 (e.g., MyMathLab and MyStatLab developed by Pearson publishing)</td>
<td>• May incur a price for students or instructors</td>
</tr>
<tr>
<td></td>
<td>• Suitable for highly motivated students</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4. Comparison of the precision of low/high at-risk levels
Depending on the nature of the individual course, different intervention methods can be employed for proactive advising. For example, Pérez (1998) suggested a number of intervention (connecting) methods for at-risk students; Jones and Hansen (2014) proposed the use of web-based synchronous tools; and Wilson and Varma-Nelson (2016) employed peer-led team learning (PLTL) for their course. Table 9 highlights the pros and cons for some commonly-used intervention methods and provides some guidelines for their deployment.

The efficacy of the above intervention methods varies with each student, and different types of intervention methods are associated with different costs. Face-to-face consultations, for instance, are costly for the instructors. Although consultations are considered an effective way for intervention, the time an instructor can afford is limited. Especially when dealing with reluctant students, it could take several hours of face-to-face consultations to change their negative attitudes to study the course. It is therefore recommended to start from the least expensive intervention method, such as emails, and then adjust the intervention strategy based on the student’s response.

After understanding the characteristics of intervention methods, some questions remain to be answered. When and with whom should the instructor intervene? How should the resources be allocated to provide effective help to the at-risk students? Should the instructor spend all his/her energy on helping the more at-risk students (which often requires greater effort) or the low-risk borderline cases? Three suggestions are made below for devising an intervention strategy which systematically guides instructors to conduct intervention for the at-risk students.

First, it should be noted that the rapport between instructors and students also affects the effectiveness of intervention (Wilson, Wilson, & Legg, 2012; Lammers & Gillaspy Jr, 2013). Instructors should therefore try to establish a good rapport with students at an early stage. For instance, sending a welcoming email to all students before the course commences, followed by personalised email reminders to every student absent from the first class. However, it is likely that not every student will respond to the emails, possibly due to overlooking or simply ignoring them. Depending on the availability of the instructor, a follow-up phone call could be made to contact the students who did not respond. As indicated in our statistical analysis, the correlation between the exam score and attendance (or clicker score) is high and significant. Therefore reminding students to contact the students who did not respond. As indicated in our statistical analysis, the correlation between the exam score and attendance (or clicker score) is high and significant. Therefore reminding students to attend the class should be adopted as one of the early intervention methods.

Second, the intervention strategy would change with stages rather than being static. In the early stages, instructors can focus more on helping the high-risk students; while, during the latter stages, most effort can be spent on the low-risk students so as to assist as many students as possible to pass the course. There would not be sufficient time to help a high-risk student to pass the course at the very last stage. Hence, it is important to identify high-risk students early to offer support.

Table 10: Systematic proactive intervention

<table>
<thead>
<tr>
<th>Stage</th>
<th>Intervention strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Send a welcoming email to all students before the course commences and ask for acknowledgement. After the first class, send a text message to all attendees and ask for acknowledgement. Ensure that students who are retaking the course and obtained low pre-requisite exam scores respond. Make follow-up phone calls if necessary.</td>
</tr>
<tr>
<td>2</td>
<td>Rank students according to the at-risk probabilities or scores provided by the prediction model. Select suitable intervention methods from the highest risk to the lowest according to the guidelines provided in Table 9. Make sure that all students respond to at least one of the communications channels.</td>
</tr>
<tr>
<td>3</td>
<td>Rank students according to the new at-risk probabilities or scores provided by the prediction model. Focus on those at-risk students who have not made significant progress thus far. Post an assignment reminder on the LMS.</td>
</tr>
<tr>
<td>4</td>
<td>Rank students according to the new at-risk probabilities or scores provided by the prediction model. Focus on high-risk students and invite them for face-to-face consultation. Post a quiz reminder on the LMS.</td>
</tr>
<tr>
<td>5</td>
<td>Rank students according to the new at-risk probabilities or scores provided by the prediction model. Focus on low-risk students and invite them for face-to-face consultation if necessary. Post an exam reminder on the LMS.</td>
</tr>
</tbody>
</table>

Third, to prioritise the intervention, the probabilities or scores provided by the prediction model can be used to rank the students according to their chance of failure. Given the possibly large number of at-risk students, such ranking is a valuable tool for the instructor to strategically allocate resources. If the instructor’s objective is to uplift academically the highest possible number of at-risk students, then he/she should concentrate on the most
responsive students (i.e., those with the highest intrinsic motivation). Table 10 summarises the systematic proactive intervention for the course in this study.

**Efficacy of intervention**

To evaluate the effectiveness of the above approach, the pass rate of the students in this study was compared with that of the whole course. This is a plausible comparison as all students in the course were taught by the same instructor and their PRE_EXAM averages were comparable (56% for the whole course; 57% for the students in the study), implying that their expected performance in the course should be similar according to the correlation test conducted in Table 2. The result indicates that the pass rate for the students in the study was 82.30%, significantly higher than 75.34% for the whole course.

Table 11 indicates the intervention success rate, which increases correspondingly with the number of interventions. The entries are statistically independent through a chi-square test, which suggests that intervention frequency is positively related to the effectiveness. In the table, occasional intervention is defined as one to two contacts with students. The majority of the cases were email reminders for attending the tutorial class, but it was found that some students never replied (and perhaps never read them). A follow-up phone call would be made if students did not show up at the next tutorial class after the email reminder. Frequent intervention means that students responded to the instructor’s email or text messages, and communicated with him/her more than twice, through various channels, including emails, phone calls, instant messaging and face-to-face consultations. This indicates that the students responded positively to previous interventions. It was found that with more contacts between instructors and students, more trust and friendship were established, as was evident from the fact that the later consultations were often requested by the students for academic questions and sometimes personal advice.

**Table 11.** Contingency table of EXAM and intervention (proportion by Intervention category in parentheses)

<table>
<thead>
<tr>
<th>EXAM</th>
<th>None (0.7894)</th>
<th>Occasionally (0.6061)</th>
<th>Frequently (1.0000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pass (Not at-risk)</td>
<td>30</td>
<td>20</td>
<td>12</td>
</tr>
<tr>
<td>Fail (At-risk)</td>
<td>8 (0.2105)</td>
<td>13 (0.3939)</td>
<td>0 (0.0000)</td>
</tr>
<tr>
<td>Chi-square: 7.8952 (p &lt; .05)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. No. of samples = 83.

In addition, Table 11 shows that eight out of 83 students (9.64%) failed the course without being noticed beforehand (i.e., false negative prediction in all five stages). Further investigation disclosed that, among the eight students who failed, two were absent from the exam due to sickness or personal reasons; and five students were within 6% below the pass grade because of careless mistakes which could actually have been avoided. The remaining failed student seemed to overlook the question requirements and made mistakes at the very beginning of two heavily weighted questions and ended up 15% below the pass grade. Identifying these unexpected cases is considered difficult and would be a trade-off between recall and precision.

It was also noted that 13 students who received occasional intervention still failed the course. A closer look at the student records revealed that four of them had encountered problems in their studies and had already decided to leave the university before the course commenced. Of the remaining nine students who failed, four gave up the exam due to the low grade they received in the assignments or quiz, and five were not very responsive to the intervention.

The students were also encouraged to form peer study groups in which they could support each other in learning. As shown in Table 12, the pass rate for students with peer groups (78.22%) is higher than those without (61.97%) and it was shown to be statistically significant according to the chi-square test.

**Table 12.** Contingency table of EXAM against group (proportion by group in parentheses)

<table>
<thead>
<tr>
<th>EXAM</th>
<th>No (0.6197)</th>
<th>Yes (0.7822)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pass (Not at-risk)</td>
<td>110</td>
<td>90</td>
</tr>
<tr>
<td>Fail (At-risk)</td>
<td>32 (0.3803)</td>
<td>11 (0.2178)</td>
</tr>
<tr>
<td>Chi-square: 5.4945 (p &lt; .05)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. No. of samples = 243.
Peer study groups are helpful not only for learning, but also for instructors’ intervention. For instance, timid students may have reservations about seeking assistance from the instructors or tutors alone, but would feel more comfortable in group consultation. Also, instructors may identify the students in the group who are more capable and encourage them to help the others in the group. There is likely to be a social multiplier effect when making interventions in peer study groups, and our preliminary result seems to support this observation. As shown in Table 13, the intervention effect on peer groups is far more successful than on individual students. The success (pass) rates of occasional interventions (i.e., “Occasionally”) on peer groups are 10/(10+2) = 83.33%, as oppose to 10/(10+11) = 47.62% on individuals. It is also worth mentioning that 11 out of 13 students who failed did not belong to any peer study group in the course. To support individual students, more intervention effort is required.

Table 13. Contingency of EXAM, Group and intervention (proportion by intervention category in parentheses)

<table>
<thead>
<tr>
<th>EXAM</th>
<th>Group</th>
<th>Intervention</th>
<th>None</th>
<th>Occasionally</th>
<th>Frequently</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>8</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.2105)</td>
<td>(0.3030)</td>
<td>(0.4167)</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td></td>
<td>22</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.5790)</td>
<td>(0.3030)</td>
<td>(0.5833)</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td></td>
<td>4</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.1053)</td>
<td>(0.3333)</td>
<td>(0.0000)</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td></td>
<td>4</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.1053)</td>
<td>(0.0606)</td>
<td>(0.0000)</td>
</tr>
</tbody>
</table>

Note. No. of samples = 83.

Discussion

This paper has presented the development of a multistage prediction model of students’ study performance and the construction of a systematic proactive intervention. In this section, the three research questions raised earlier are discussed.

For research question 1, the prediction results have confirmed the effectiveness of using clicker data for identifying students at risk. It was found that QUIZ, PRE_EXAM, and ASSIGNMENT are respectively the three most important predictive factors, followed by CLICKER. Among all the predictive variables, INTEREST and EXPECT become insignificant in predicting the outcome of the final exam after stage 2. This observation is consistent with the finding of Tempelaar et al. (2015), who reported that the predictive power of learning disposition data is dominated by assessments. While learning disposition data seem to be of little value in improving prediction performance, the information is still valuable at the early stage, especially when formative and summative assessments are not yet available. In addition, a student’s intrinsic and extrinsic motivations offer insightful information for instructors to choose an appropriate intervention method. For example, students who have a great interest in mathematics may be more motivated in self-learning and therefore are more suitable for student-centric intervention methods such as e-tutorials. Students who have a low expectation for their grade may need academic counselling to stimulate their attitudes, behaviour and cognition (Rienties et al., 2017).

For research question 2, this paper proposed employing LR with prediction intervals instead of LG for predicting the students who would potentially fail. There has been discussion on whether LR should be used for dichotomy classification (Hellevik, 2009). In this study, we argue that using LR with prediction intervals is reasonable as we are regressing students’ exam scores instead of directly predicting the pass and fail grade. In such a case, LR makes use of more information (exam score) for regression, while LG ignores the exam scores and maps it into two categories (pass and fail) for classification. Empirically, LR with prediction intervals has shown a better performance than LG in terms of recall and is comparable to LG in precision. As identifying potential fail students is essential, recall is regarded as more important than precision.

For research question 3, the result confirmed the effectiveness of the systematic proactive intervention. The effectiveness of the proposed strategy with intervention frequency showed that students who received more interventions were more likely to pass the exam. This observation is consistent with the quantitative study on the relationship between academic advising and study retention conducted by Swecker, Fifolt and Searby (2013). Peer study groups have also proved to be beneficial in terms of passing the exam. This is consistent with the findings of relevant studies such as Boud, Cohen, and Sampson (2014).
Conclusions

This study contributes to showing how in-class clicker data can be used to identify students at risk of failing a course. The clicker data were collected using available free online tools and students’ own mobile devices which serve as an LA solution at a low cost. The multi-stage prediction model showed that the types of data vary in effectiveness in identifying at-risk students at different stages of a course. For the choice of the regression model, this study found that LR with prediction intervals, which was less frequently used than LG in previous studies, could be more suitable for its better performance in recall.

The efficacy of intervention methods and their effective use have also been analysed. They contribute to closing the loop in the LA cycle, which has been recognised as a major challenge ahead of LA research and practice (Clow, 2012; Rienties et al., 2017). The strategic implementation of proactive intervention in relation to the risk levels of students and different stages of a course is a potential way to cope with the limited time and resources of instructors, while helping at-risk students to pass the course to a greatest extent.

Future work

In the case study reported, students received their clicker scores via emails along with the correct answers after the tutorial class. It would be more desirable to use a dashboard for students to examine at any time whether they are at risk. In addition to identifying at-risk students, our work can be extended to measure the probability of students achieving their desired grades in the course; and this can be done through linear regression with the prediction intervals proposed in this paper or ordinal logistic regression analysis (O’Connell, 2006), so that students can be informed of their continual learning progress and how likely it is that they can achieve the grades they want in the course.

The intervention strategy proposed was largely based on the instructors’ own teaching experience and discretion. The adoption of a proper LA intervention design (Wise, 2014) is therefore crucial. Also, in order to systematically design an effective intervention strategy, a more rigorous and objective evaluation of the effectiveness of interventions is necessary. For instance, the instruments proposed by Moody and Sindre (2003) and Muñoz-Merino, Ruipérez-Valiente, Alario-Hoyos, Pérez-Sanagustín and Kloos (2015) can be adopted to measure quantitatively the effectiveness of the learning interventions on various types of students. In addition, an evidence-based research approach (Rienties et al., 2016) can be adopted to further validate the findings presented in this paper. After the quantitative measures of the impact of intervention become available, the reinforcement learning framework (Choi & Lam, 2018) can then be applied to take intervention effects into consideration so as to build a sophisticated decision model for intervention strategies.

Although peer instruction has been proposed for decades (Mazur, 1997), it has now become a popular pedagogy for engaging students in class discussion. Peer instruction can also be adopted to enhance the use of clicker through the following process. After the instructor poses the question, students are first asked to provide their own answer by using clickers. The instructor shows the distribution of the answers provided and asks the students to discuss their answers with their classmates; and, after the discussion, the students are given a chance to change their answers. The instructor then displays the new statistics for the answers and leads a class discussion. Prior research has reported that peer instruction has a positive learning impact in physics and computer science courses. It would be interesting to examine whether this pedagogic technique also applies to business quantitative methods courses.

Acknowledgements

The work described in this paper was partially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (UGC/IDS16/15).

References


Hellevik, O. (2009). Linear versus logistic regression when the dependent variable is a dichotomy. *Quality & Quantity, 43*(1), 59–74.


A Peer Coaching-based Professional Development Approach to Improving the Learning Participation and Learning Design Skills of In-Service Teachers

Ning Ma1*, Shuang Xin2 and Jia-Yuan Du2

1Faculty of Education, Beijing Normal University, Beijing, China // 2Institute of Education, University College London // horsening@bnu.edu.cn // heart_shuang@163.com // rtnvjdu@ucl.ac.uk

ABSTRACT

Personalized learning based on learning analytics has become increasingly important for teachers’ development via providing adaptive contents and strategies for teachers by identifying their questions and needs. Currently, most studies on teachers’ professional development focus on pre-service teachers, and studies on teachers’ personalized learning focus on the expert guidance approach. In this paper, a peer coaching-based personalized learning approach is proposed to help in-service teachers identify their questions and needs and adapt their teaching plans based on peer feedback as a result of interacting with their peers and reflecting on their work so as to engage in in-depth learning and transfer of knowledge to their teaching practice. In order to evaluate the effectiveness of the proposed approach, a quasi-experimental design was employed, involving 20 in-service Mandarin teachers. The experimental group teachers learned with the peer coaching-based personalized learning approach, while the control group teachers learned with the expert guidance-based personalized learning approach. The study was conducted using a quantitative approach. The instruments used were a learning participation rubric and performance assessments of the participating teachers’ lesson plans and teaching videos. The findings indicated that the post-test scores of the experimental group were significantly higher than those of the control group. The peer coaching-based personalized learning approach had a much better effect than the expert guidance-based personalized learning approach on the in-service teachers’ learning participation, learning design skills, and in-practice teaching abilities.

Keywords

Personalized learning, Peer coaching, Learning participation, Learning design skill, Learning analytics, Teaching ability

Introduction

The development of the prospective teachers’ knowledge, skills, and dispositions is a key element for high-performing countries’ success (Darling-Hammond et al., 2017). Within an education culture craving continuous improvement, schools and society have constant needs to ensure that teachers’ skills, knowledge, and actions match the changing environment (Lindon, 2011). Teachers have a great influence on students’ performance throughout a large span of their school careers. They can help cultivate students’ habits of mind and knowledge schemes, thereby enabling them to make meaningful contributions and to prosper in the open, technological world of the future (Darling-Hammond, 2000). Improving schools, enhancing teaching quality, and improving the quality of students’ study are so important that it has led to a focus on Professional Development for Teachers as an important way to achieve these goals (Opfer & Pedder, 2011). Official documents in China and reports from international institutions (Lo, Lai, & Wang, 2013; van den Bergh, Ros, & Beijaard, 2015) regard Professional Development for Teachers as an important factor in educational improvement.

To improve education quality and promote teachers’ professional development, education institutions and governments at different levels have taken various measures and actions (UNESCO, 2015). Personalized learning based on learning analytics has been further pointed out by scholars as helping teachers improve their lesson plans and learning material design (Ganser, 2000). Learning analytics provides helpful suggestions to instructors and learners by analyzing learning information or educational data (Hwang, Hung, Chen, & Liu, 2014). One of the objectives of learning analytics is to identify learners’ learning status or problems by analyzing their learning behaviors or interactive content, and providing adaptive and personalized learning contents, user interfaces, or practices (Hwang, Chu, & Yin, 2017). In recent years, personalized learning based on learning analytics has attracted much attention from the education field due to its characteristics of respecting the differences of individuals, emphasizing trainees’ status as the subject, and the abilities to solve teachers’ personalized problems (Wongsopawiro, Zwart, & van Driel, 2017). Moreover, several explorations related to personalized learning in the online context have been made, such as studies on personalized e-learning platform construction for elementary and secondary school students (Capuano et al., 2014; McLoughlin & Lee, 2010; Peter, Bacon, & Dastbaz, 2010), personalized vocational training frameworks (Mellett & O’Brien, 2014), and online personalized teacher training modes based on diagnosis of the teaching design (Li & Ma, 2014).
Holly (1989) indicated that teachers’ perceptions of professionalism are mainly gained from “other teachers.” This can be achieved through peer coaching activities. In this research, we therefore combined peer coaching with teachers’ personalized learning, with the aim of exploring the effects of peer coaching during teachers’ online personalized learning processes. However, an overview of the global research on teachers’ personalized learning also revealed several common phenomena, such as the emphasis on trainer-and-trainee interactions rather than on peer interactions (Rangel et al., 2015; Steiner, Dobbins, & Trahan, 1991), the difficulty trainers face in providing personalized suggestions to individual trainees (Dennis et al., 2018), and the poor teaching outcomes (Atueyi, 2016). That is, most previous studies related to personalized teacher development have attempted to identify the learning status or problems of the trainees and to provide recommendations or contents for them from the perspectives of the experts, while little research has been conducted which comprehensively integrates learning analytics and personalized learning from the perspectives of peers.

This research used an online platform in a 5-week quasi-experiment to find out how the peer coaching-based personalized learning approach would help enhance in-service teachers’ learning participation and affect their teaching design skills and teaching abilities in practice compared to the expert guidance-based personalized learning approach. We hope to provide policy-makers, instructors, and teachers with alternative, more effective approaches to teachers’ future professional development.

**Literature review**

**Professional development for teachers**

Teacher education has become an essential area of government policy in many countries around the world over the last 30 years (Furlong, 2013); in particular, teacher preparation has been recognized as an important and challenging issue for most public universities in many countries (AASCU, 2016). Professional Development for Teachers is “about teachers learning, learning how to learn, and transforming their knowledge into practice for the benefit of their students’ growth” (Avalos, 2011). In other words, it organizes learning to improve teachers’ professional skills and knowledge of students’ performance (Hill, Beisiegel, & Jacob, 2013). Professional Development for Teachers can help them build their knowledge and beliefs, address perceived problems, and develop their classroom practices (Opfer & Pedder, 2011). It can provide teachers with opportunities to develop expertise in the curriculum, instruction, and the assessment of student learning, finally resulting in improvements in students’ educational outcomes (Tait-McCutcheon & Drake, 2016).

A variety of methods have been used to improve teachers’ professional development. The most traditionally used method was to invite experts to teach knowledge or practices using a face-to-face approach (Zhang, Liu, & Wang, 2017). With the rapid development of internet technology, however, the method has changed from traditional face-to-face training to training in advanced online environments (Chen, Chen, & Tsai, 2009). This new approach allows trainees to interact in the online environment at anytime and anywhere that is convenient (Al-Balushi, & Al-Abdali, 2015). Online training plays a significant role in teachers’ professional development (Jimenez & O’Sahanahan, 2016; Kao, Tsai, & Shih, 2014), especially online personalized learning, which can provide teachers with personalized courses and materials, can adopt to their learning styles and progress, and allows them to take advantage of the online environment (Limongelli, Sciarrone, Temperini, & Vaste, 2011). This kind of learning has been found to have a positive influence on teachers’ professional development (Gynther, 2016).

After a 3-month design-based study, Li and Ma (2014) built an expert guidance-based online personalized learning model for teachers’ development which included three stages: diagnosis, personalized recommendation, and personalized evaluation. They also found that expert guidance-based personalized learning could promote the teachers’ learning design skills. However, their research also found that the teachers did not benefit so much in terms of some high-level skills, such as applying the pedagogies in practice. Li and Ma (2014) argued that the personalized diagnosis, adaptive learning, and the interactions between the trainers and trainees may have promoted the development of the trainees’ knowledge and skills, but the weak interactions between the trainees could be an important factor affecting their in-depth learning and the development of their advanced skills.

Teachers who know the teaching contents or pedagogies may not be able to apply them in their teaching practice. Expert knowledge and understandings of pedagogies are prerequisites but not guarantees that teachers will teach well. It also does not mean that they know what concepts are difficult for students, what representations are best for certain ideas, or what ways are optimal for developing conceptual understandings (Lindon, 2011). Peer
coaching might thus be an alternative powerful approach for teachers’ professional improvement (Rice, 2012; Zhang, Liu, & Wang, 2017).

Peer coaching

Peer coaching generally involves two colleagues engaged in a mutually supportive relationship (Neubert & McAllister, 1993). It is a confidential process through which instructors provide one another with assistance, feedback, and support, and share their expertise, for the purpose of enhancing learning (Kohler et al., 1997). Many studies have emphasized the importance of teachers as the subjects of peer coaching (Alsaleh et al., 2017; Yu, 2003; Zhang, Liu, & Wang, 2017). For example, Alsaleh’s (2017) study shared that the peer coaching enhanced teachers’ professional development based on teaching practices, teacher learning, team cooperation, and teachers’ self-confidence, enthusiasm, and autonomy. Meanwhile, many studies have reported the effectiveness of a learning approach and environment that involves peer coaching, peer assessment, and peer review in different fields (Hsu, 2016; Papadopoulos, Lagkas, & Demetriadis, 2017; Yu & Wu, 2016).

In the past decades, peer coaching has been one strategy espoused by teacher education programs around the world to enhance the experience and development of teachers, and has also been evidenced in the literature as being helpful in various aspects of field-based experience (Lu, 2010). For instance, Goker (2006) implemented peer coaching in pre-service TEFL teacher education. He found that the student teachers’ instructional skills and self-efficacy were significantly improved compared to those just receiving traditional supervisor visits. Some studies have reported the effectiveness of reflection in peer coaching and peer assessment strategies. A study on technology-enhanced peer reviews provided evidence that the review “giver” perspective is a vital option for peer reviews (Papadopoulos, Lagkas, & Demetriadis, 2017). Hwang, Hung, and Chen (2014) reported the effectiveness of adopting the peer-assessment approach in terms of helping students make reflections on and improve their digital storytelling projects. Peer coaching enriches teachers’ reflections on their practices, and thus enhances and invigorates teachers’ teaching skills.

Meanwhile, more studies have been reported involving pre-service teachers than in-service teachers in the peer coaching and peer assessment field (Lu, 2010). Peer coaching in pre-service teacher education has its unique advantages, such as the similar experience and knowledge levels of the student teachers, the same courses or time frame that the student teachers are engaged in, as well as the cost efficiency in the program curriculum (Lu, 2010). All these advantages that could sustain the feasibility and serve as a rationale for the incorporation of peer coaching in pre-service teacher education, also imply the possible challenges and obstacles of in-service teacher education. Till now, there has been very little scholarship regarding whether peer coaching could be implemented in a regular teacher training program.

Therefore, in this study, a peer coaching-based personalized learning approach is proposed for in-service teachers. A quasi-experiment was also conducted to investigate the effectiveness of the proposed approach regarding the development of the teachers’ learning participation, learning design skills, and their in-practice teaching abilities.

Research questions

In this study, a peer coaching-based personalized learning approach is proposed for in-service teachers. It was expected that the proposed approach could benefit in-service teachers in terms of improving their online learning participation and promoting their learning design skills and their in-practice teaching abilities. Accordingly, the following research questions were investigated:

- Does the peer coaching-based personalized learning approach benefit in-service teachers more than the expert guidance-based personalized learning approach in terms of online learning participation?
- Can the peer coaching-based personalized learning approach promote in-service teachers’ learning design skills in comparison with the expert guidance-based personalized learning approach?
- Does the peer coaching-based personalized learning approach benefit in-service teachers more than the expert guidance-based personalized learning approach concerning the advanced abilities to apply the teaching knowledge and skills in practice?
Peer coaching-based personalized learning system for in-service teachers

In this section, the personalized learning system which can support the expert guidance approach and peer coaching approach is demonstrated.

Online personalized learning system for in-service teachers

By referring to the online personalized learning model for teachers established by Li and Ma (2014), an online personalized learning system for in-service teachers was developed (see Figure 1) in this study. The system contained four main modules, namely object analysis, personalized diagnosis, personalized recommendation, and personalized evaluation. In the first stage, trainees enter the system and complete the surveys, such as their basic information and ICT literacy. Each trainee needs to submit a lesson plan which is used for the object analysis and the next stage. In the second stage, through analyzing each trainee’s lesson plan based on the diagnosis framework, problems are identified for each trainee. Then, personalized learning contents and activities are recommended for each trainee in the third stage. In this stage, the trainees can learn individually online following their own personalized learning contents and activities. The last stage is the evaluation and feedback stage for the trainees. They could make self-reflections at this stage.

For the expert guidance-based personalized learning approach, the trainees can get support from the experts throughout the whole learning process. For the peer coaching-based personalized learning approach, the trainees can interact with their peers and get support from each other.

Peer coaching activities in the personalized learning system

According to the social cognitive theory, Stahl (2000) divided the knowledge-building process into two parts, namely personal understanding and social knowledge building; these two aspects contribute to the development of each other. Stahl (2004) pointed out that a system or an activity to support collaborative knowledge building should include functions that can support collaboration, social awareness, knowledge building, and knowledge management. Swafford (1998) also argued that peer coaching activities for teachers should provide chances for them to discuss, analyze, and reflect on their classroom instruction. On the basis of the above studies, this research adopted the following principles to design the peer coaching activities: (1) providing the teachers with opportunities to face specific teaching problems posed by their peers and by themselves; (2) promoting the teachers’ experience by sharing among peers; (3) providing opportunities for the teachers to solve problems among themselves; and (4) promoting the teachers’ real-time reflection. Following these principles, we organically integrated the peer coaching activities in the process of object analysis, personalized diagnosis, personalized recommendation, and personalized evaluation in teachers’ online personalized learning, as shown in Figure 2.
In this system, the peer coaching activities during the online personalized learning process consist of the following aspects.

(1) Meeting and greeting virtual team members. Building trust is the basis for peer coaching. At the beginning, the teachers do not know each other in the cyberspace. By greeting each other, posting their photos and other related information, they can get to know each other, bridge the gap, and build a sense of belonging and trust. This will help to develop the follow-up peer coaching activities.

(2) General content learning and peer diagnosis. Before the peer diagnosis, learning about the general theories, pedagogies, and the usage of the Instructional Design Diagnosis Framework is very important. After that, teachers can review the lesson plans in groups, diagnose problems, and give some feedback to each other according to the diagnosis framework. They can download the lesson plans, check the diagnosis framework, learn the general content, give suggestions, take notes, and so on, as shown in Figure 3. Based on the diagnoses and suggestions teachers give each other, the system can build a recommended learning contents and activities list for every teacher.

(3) Personalized content learning and peer experience sharing. After the peer diagnosis, every teacher was given a personalized learning contents and activities list. They could learn individually and share their experiences during this learning stage.

(4) Self-revision. The teachers could review and revise their pre-lesson plans by themselves based on the knowledge they gained through the personalized learning and peer experience sharing.

(5) Peer revision. After the self-revision of the pre-lesson plans, the teachers could share their plans in groups again and give each other comments. An online tool with the function of sharing documents and joint editing was provided for the teachers. They could review and make comments on every lesson plan, check others’ comments, discuss and negotiate regarding the comments, rethink the knowledge that they had learned, and accumulate practical experience of how to apply the knowledge in practice.

(6) Raising questions and exchanging ideas. The system provides an interaction tool for teachers to raise questions and to communicate throughout the whole learning process. It also encourages the teachers to give explanations, reply to the questions, and engage in deep discussion with a function for making notes and joint editing online, as shown in Figure 4. That is, teachers raise questions first. Then, they can discuss with their peers and help each other to solve the problems. In this case, the trainee acts as both a participant and a training expert, correcting the imprecise or inaccurate opinions, and also playing the role of an expert in the online personalized learning.

Figure 2. Peer coaching activities in the personalized learning system design
Figure 3. Interface of diagnosis in peers

Figure 4. Interface of raising questions and exchanging ideas
Methodology

Based on the above personalized learning system, a quasi-experimental design was conducted involving in-service Mandarin teachers. The objectives of the course were to foster the teachers’ learning design skills and their abilities of applying their knowledge and skills in their teaching practice.

Participants

The participants were 20 in-service teachers (all females) who had taught the Mandarin course for 7.16 years on average in elementary schools. The average age of the teachers was 31.56 years old. All the participants had previous experience of online learning.

Learning activities and experimental procedure

A quasi-experimental design was used to compare the learning participation, learning design skills, and in-practice teaching abilities of the in-service teachers who learned with the peer coaching-based personalized learning approach and those who learned with the expert guidance-based personalized learning approach. Figure 5 shows the procedure of the experiment.

At the beginning of the experiment, a questionnaire about the in-service teachers’ basic information, such as their age, number of years teaching, grade that they teach, and experience of online learning was conducted. Every teacher was asked to write and submit a lesson plan according to a specific topic. According to the quality of the lesson plan and their basic information, the participants were divided into two groups at the same initial level. Using purposive sampling, we randomly selected one group as the experimental group and the other as the control group.

Teachers in both groups took part in a 5-week personalized learning program based on the problems reflected in their initial lesson plans. The difference between the two groups was that the experimental group used the peer coaching-based personalized learning system, while the control group used the expert guidance-based personalized learning approach. Analysis of learning behaviors and interaction data, post-lesson plan, teaching videos in the classroom.

Figure 5. Procedure of the experiment
coaching-based personalized learning approach, while the control group used the expert guidance-based personalized learning approach.

When the control group teachers entered the learning platform and submitted their initial lesson plan, the training experts diagnosed the problems existing in the plans according to the Instructional Design Diagnosis Framework, and then recommended a personalized learning contents and activities list for each trainee. The participants could check the course list and learn independently and adaptively. When the trainees met questions or problems, they could post their questions in the discussion forum, and the training experts would give them feedback. The experts also gave some learning suggestions, additional learning materials, or raised some questions according to each teacher’s individual needs. For example, if a trainee did not raise any questions or ask for any information, the experts would communicate with her and give some suggestions or try to provide some support for her.

When the experimental group teachers first entered the learning platform, they introduced themselves and got to know each other. Then, they checked their peers’ lesson plans and performed diagnosis on them according to the Instructional Design Diagnosis Framework, proposed personalized course lists for each trainee, and recommended the lists to them. Trainees in the experimental group studied individually referring to the lists, but during the process, they could communicate with their peers about any questions or problems that they met or found.

After 5 weeks of learning, a post-test was carried out to check and evaluate the differences in the teachers’ learning participation, learning design skills, and in-practice teaching abilities. These results were mainly derived from the analysis of the online learning process, the quality of the post-lesson plans, and videos of the participants teaching in the classroom after the learning process.

**Instruments**

To evaluate the effectiveness of the proposed approach, an evaluation framework for lesson plans, a coding scheme for learning participation analysis, and an evaluation framework for the teaching videos were employed for the pre- and post-test in the experiment.

**Table 1. The instructional design diagnosis framework**

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Assessment items</th>
<th>Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Front-end analysis</td>
<td>Analyse the learners</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Identify and describe the learning goals</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Identify and describe the learning content, especially the important and difficult learning points</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Describe the core pedagogy or teaching ideas</td>
<td>4</td>
</tr>
<tr>
<td>Learning process design</td>
<td>Design an appropriate learning context and lead into the learning fluently</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Set or post appropriate tasks/questions/problems</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Design rich, interesting, and effective learning activities that can promote the students’ learning attitude and deep learning</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Select or develop effective strategies for guiding the students’ reading, such as role play, teacher modelling, etc.</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Select or develop strategies to help the students grasp the method of literacy learning, especially for writing</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Select or develop rich materials for extensive reading that focused on the learning goals</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Set effective and proper writing items, provide relevant scaffolding for writing</td>
<td>8</td>
</tr>
<tr>
<td>Pedagogies and teaching ideas</td>
<td>The 211 teaching approach (20 minutes for studying the textbook, 10 minutes for extensive reading, and 10 minutes for composition writing)</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Appropriate integration of literacy, reading and writing</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>The role of the teacher and the students (students as the principal part of the learning, teachers as the assistant and supporter of the students)</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Development of the students’ creative thinking in language learning</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Development of the students’ critical thinking in language learning</td>
<td>8</td>
</tr>
</tbody>
</table>

For the diagnosis and assessment of the teachers’ pre- and post-lesson plans, the Instructional Design Diagnosis Framework was modified from the measurement developed by Li and Ma (2014). The framework was developed
by three experienced experts who had more than 10 years’ experience of Mandarin teaching. It consists of three dimensions and 16 assessment items, as shown in Table 1. Using the Analytic Hierarchy Process (AHP) approach, the researchers checked the consistency of the weighted scores given by the three experts to derive the scores for each item.

Before and after the experiment, three experts evaluated the pre- and post-lesson plans of the in-service teachers according to the framework. Kendall’s coefficient of concordance was used to test the consistency of the results for the three aspects, with Kendall’s W of 0.83 (p = .000) and 0.72 (p = .002), respectively. Therefore, the pre- and post-test scores given by the three experts showed significant consistency, and the scores were valid.

For the assessment of the online learning participation, the learning behaviours and the interaction information of the teachers recorded by the personalized learning system were analyzed. In this research, all the participants learned online. The experimental group participants interacted with their peers, while the control group interacted with the expert. Seven dimensions were selected from the online knowledge-building rubric proposed by Li and Ma (2011) as the learning participation instrument to code and count the participants’ learning behaviours and interaction data. It consisted of seven dimensions, namely raising questions, discovery and explanation, conflict, support, reflection, sharing, and affective communication, with a total of 19 items, as shown in Table 2. The learning behaviours and the interaction data of the participants were coded by the instrument, and the numbers of each dimension were counted. The scores of each dimension were the count numbers. Then, an independent t-test was conducted on the average numbers of the seven dimensions for the two groups.

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Assessment items</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raising questions</td>
<td>Asking for information</td>
<td>1a</td>
</tr>
<tr>
<td></td>
<td>Raising a well-structured question</td>
<td>1b</td>
</tr>
<tr>
<td></td>
<td>Raising an ill-structured question</td>
<td>1c</td>
</tr>
<tr>
<td>Discovery and explanation</td>
<td>Briefly statement about the concepts, definitions and facts.</td>
<td>2a</td>
</tr>
<tr>
<td></td>
<td>Building relationship between the facts, ideas and principles</td>
<td>2b</td>
</tr>
<tr>
<td></td>
<td>Identify or analyzing a question</td>
<td>2c</td>
</tr>
<tr>
<td></td>
<td>Clarifying a question through analogizing</td>
<td>2d</td>
</tr>
<tr>
<td></td>
<td>Clarifying a question through comparing</td>
<td>2e</td>
</tr>
<tr>
<td></td>
<td>Distinguish the reason and outcome, advantage and disadvantage</td>
<td>2f</td>
</tr>
<tr>
<td></td>
<td>Proving a point</td>
<td>2g</td>
</tr>
<tr>
<td>Conflict</td>
<td>Strongly opposed to a view</td>
<td>3a</td>
</tr>
<tr>
<td></td>
<td>Raising a disagreement to a view</td>
<td>3b</td>
</tr>
<tr>
<td>Support</td>
<td>Agreement to a view</td>
<td>4</td>
</tr>
<tr>
<td>Reflection</td>
<td>Summarizing the learning outcomes</td>
<td>5a</td>
</tr>
<tr>
<td></td>
<td>Self-reflection</td>
<td>5b</td>
</tr>
<tr>
<td>Sharing</td>
<td>Sharing a view, solution, information, web site, material, etc.</td>
<td>6</td>
</tr>
<tr>
<td>Affect</td>
<td>Greeting to others</td>
<td>7a</td>
</tr>
<tr>
<td>communication</td>
<td>Expression of friendship, encouragement, support, understanding, funny, agreement, etc.</td>
<td>7b</td>
</tr>
<tr>
<td></td>
<td>Expression of rejection, depression, worry, etc.</td>
<td>7c</td>
</tr>
</tbody>
</table>

In order to assess the teachers’ in-practice teaching abilities, each participant submitted a 40-minute teaching video showing how they applied the knowledge, skills, and the lesson plan in the classroom after the learning process. The last two dimensions of the “Instructional Design Diagnosis Framework” were used as the assessment tool, which were the dimensions of learning process design and pedagogies and teaching ideas. Although the names of the dimensions and the relevant items of these two assessment tools are the same, the concerns are different. The lesson plan shows the ideas and plans of the teacher, while the teaching video shows the actual teaching abilities reflected in the teacher’s behaviours and activities. A teacher who can write a good lesson plan is not necessarily a teacher who can teach well in an actual classroom. Three experts evaluated the teaching videos of the in-service teachers according to the frameworks. Kendall’s coefficient of concordance was used to test the consistency of the results for the three aspects, with Kendall’s W of 0.78 (p = .001). Therefore, the evaluated scores of the teaching videos given by the three experts showed significant consistency, and the scores were valid.
Research results

Learning participation

To understand if there was a difference in the learning participation of the experimental group and control group teachers, content analysis was conducted firstly on the participation information in the online communities of the two groups. Then, an independent t-test was conducted on the seven dimensions of the teachers’ learning participation. As shown in Table 3, there was a significant difference between the two groups in all seven dimensions. The experimental group showed a significantly higher occurrence of raising questions ($t = 2.11$, $p < .05$, Cohen’s $d = 1.05$), discovery and explanation ($t = 5.02$, $p < .01$, Cohen’s $d = 2.25$), conflict ($t = 4.16$, $p < .01$, Cohen’s $d = 1.86$), support ($t = 3.04$, $p < .05$, Cohen’s $d = 1.36$), reflection ($t = 9.00$, $p < .001$, Cohen’s $d = 3.98$), sharing ($t = 9.51$, $p < .001$, Cohen’s $d = 4.24$) and affective communication ($t = 2.67$, $p < .05$, Cohen’s $d = 1.19$) than the control group. Furthermore, Cohen (1988) indicated that a Cohen’s $d$ value greater than 0.50 represents a medium effect size, while a Cohen’s $d$ value greater than 0.80 represents a large effect size; this result, therefore indicated a rather good effect size.

Table 3. Summary of the t-test analysis of the seven dimensions of the teachers’ learning participation

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Experimental group</th>
<th>Control group</th>
<th>$t$</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raising questions</td>
<td>10</td>
<td>10</td>
<td>2.11</td>
<td>1.05</td>
</tr>
<tr>
<td>Discovery and explanation</td>
<td>10</td>
<td>10</td>
<td>5.02</td>
<td>2.25</td>
</tr>
<tr>
<td>Conflict</td>
<td>10</td>
<td>10</td>
<td>4.16</td>
<td>1.86</td>
</tr>
<tr>
<td>Support</td>
<td>10</td>
<td>10</td>
<td>3.04</td>
<td>1.36</td>
</tr>
<tr>
<td>Reflection</td>
<td>10</td>
<td>10</td>
<td>9.00</td>
<td>3.98</td>
</tr>
<tr>
<td>Sharing</td>
<td>10</td>
<td>10</td>
<td>9.51</td>
<td>4.24</td>
</tr>
<tr>
<td>Affective communication</td>
<td>10</td>
<td>10</td>
<td>2.67</td>
<td>1.19</td>
</tr>
</tbody>
</table>

Note. ’$p < .05$; **$p < .01$; ***$p < .001$.

Learning design skills

The total scores for the learning design skills include three dimensions: front-end analysis, learning process design, and pedagogies and teaching ideas. Before the analysis of the learning design skills based on the lesson plans, an independent t-test was used to analyse the pre-test. Table 4 shows that these two groups did not significantly differ in their scores for the four aspects before the experiment.

Table 4. Summary of the t-test analysis of the pre-test scores for learning design skills

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Experimental group</th>
<th>Control group</th>
<th>$t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total scores</td>
<td>10</td>
<td>10</td>
<td>-0.22</td>
</tr>
<tr>
<td>Front-end analysis</td>
<td>10</td>
<td>10</td>
<td>0.49</td>
</tr>
<tr>
<td>Learning process design</td>
<td>10</td>
<td>10</td>
<td>-0.64</td>
</tr>
<tr>
<td>Pedagogies and teaching ideas</td>
<td>10</td>
<td>10</td>
<td>-0.12</td>
</tr>
</tbody>
</table>

One-way ANCOVA was then used to compare the scores of the front-end analysis, learning process design, pedagogies and teaching ideas, and the total scores of the post-test for the two groups. The results were shown in Table 5. It was found that the teachers in the experimental group had significantly higher total scores than those in the control group for their total scores ($F = 22.31$, $p < .001$, $\eta^2 = 0.57$), front-end analysis ($F = 4.79$, $p < .05$, $\eta^2 = 0.22$), learning process design ($F = 24.21$, $p < .001$, $\eta^2 = 0.59$), and pedagogies and teaching ideas ($F = 24.56$, $p < .001$, $\eta^2 = 0.59$). The data analysis above showed that there was a significant difference between the
experimental group and the control group in terms of their learning design skills, and these differences were mainly reflected in the learning process design, pedagogies and teaching ideas, and total scores.

Table 5. The one-way ANCOVA result of the post-lesson plan scores of the two groups

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Adjusted mean</th>
<th>Std. error</th>
<th>F value</th>
<th>η²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental group</td>
<td>10</td>
<td>78.57</td>
<td>4.26</td>
<td>78.82</td>
<td>1.81</td>
<td>22.31***</td>
<td>0.57</td>
</tr>
<tr>
<td>Control group</td>
<td>10</td>
<td>66.95</td>
<td>9.85</td>
<td>66.70</td>
<td>1.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Front-end analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental group</td>
<td>10</td>
<td>11.34</td>
<td>0.82</td>
<td>11.22</td>
<td>0.54</td>
<td>4.79</td>
<td>0.22</td>
</tr>
<tr>
<td>Control group</td>
<td>10</td>
<td>9.45</td>
<td>2.60</td>
<td>9.56</td>
<td>0.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning process design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental group</td>
<td>10</td>
<td>45.22</td>
<td>2.71</td>
<td>45.57</td>
<td>0.95</td>
<td>24.21***</td>
<td>0.59</td>
</tr>
<tr>
<td>Control group</td>
<td>10</td>
<td>39.28</td>
<td>4.62</td>
<td>38.94</td>
<td>0.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pedagogies and teaching ideas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental group</td>
<td>10</td>
<td>22.02</td>
<td>1.37</td>
<td>22.07</td>
<td>0.56</td>
<td>24.56***</td>
<td>0.59</td>
</tr>
<tr>
<td>Control group</td>
<td>10</td>
<td>18.22</td>
<td>3.25</td>
<td>18.17</td>
<td>0.56</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .05; **p < .01; ***p < .001.

Teaching abilities in practice

An independent t-test was conducted on the two groups to analyse the teaching videos submitted by the participants. The results given in Table 6 indicated that there were significant differences between the two groups in terms of their total scores ($t = 4.27$, $p < .01$, Cohen’s $d = 1.91$), learning process design ($t = 4.09$, $p < .01$, Cohen’s $d = 1.83$), and pedagogies and teaching ideas ($t = 4.17$, $p < .01$, Cohen’s $d = 1.86$). According to the Cohen’s $d$ value, this result indicated a rather good effect size.

Table 6. Summary of the t-test analysis of the teaching videos of the two groups

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>$t$</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total scores</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental group</td>
<td>10</td>
<td>63.77</td>
<td>1.53</td>
<td>4.27**</td>
<td>1.91</td>
</tr>
<tr>
<td>Control group</td>
<td>10</td>
<td>54.43</td>
<td>6.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning process design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental group</td>
<td>10</td>
<td>40.70</td>
<td>0.98</td>
<td>4.09**</td>
<td>1.83</td>
</tr>
<tr>
<td>Control group</td>
<td>10</td>
<td>35.27</td>
<td>4.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pedagogies and teaching ideas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental group</td>
<td>10</td>
<td>23.07</td>
<td>0.91</td>
<td>4.17**</td>
<td>1.86</td>
</tr>
<tr>
<td>Control group</td>
<td>10</td>
<td>19.17</td>
<td>2.82</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. **$p < .01$.

Summary and discussion

Recent research has shown that personalized learning for teachers has a positive influence on teachers’ professional development (Gynther, 2016; Limongelli, Sciaronne, Temperini, & Vaste, 2011). Studies have also indicated that peer coaching is a powerful approach for teachers’ professional improvement, as teachers’ perceptions of professionalism are mainly gained from “other teachers” and so peer coaching can help teachers transform their knowledge into practice (Zhang, Liu, & Wang, 2017; Rice, 2012). However, more studies have reported on pre-service teachers than on in-service teachers (Lu, 2010). In this study, we propose a peer coaching-based personalized learning approach for in-service teachers compared to the expert guidance-based personalized learning approach. A 5-week quasi-experiment was conducted to investigate the influences on the teachers’ learning participation and the development of the teachers’ learning design skills and in-practice teaching abilities.

The peer coaching-based personalized learning approach promotes in-service teachers’ learning participation

The experimental results showed that the peer coaching-based personalized learning approach for in-service teachers could improve their learning participation. Peer coaching made the interaction between the trainees and trainers no longer the only important interactive form in the in-service teachers’ personalized learning. Peer diagnosis, raising questions and exchanging ideas, and other activities could promote the interaction between the trainees, stimulating the in-service teachers to ask more questions. Because of the similarities in their experience and background, it might be easier for the in-service teachers to express and accept their peers’ comments compared with learning with expert guidance. Besides, the in-service teachers could propose solutions to questions based on their own experience, which could give other in-service teachers a better reference.
Peer coaching benefits the development of in-service teachers’ learning design skills and in-practice teaching abilities

The research results showed that the peer coaching-based personalized learning approach had a significant influence on the teachers’ learning design skills and in-practice teaching abilities. The significant differences in the learning design skills of the experimental and control groups were mainly reflected in two dimensions: the learning process design, and the pedagogies and teaching ideas. These two parts both focus on in-service teachers’ knowledge structure based on previous experience, and transformation of theoretical knowledge into practical knowledge. The peer coaching activities in the personalized learning process helped to enrich the in-service teachers’ practical knowledge and teaching context. They also benefited the in-service teachers in terms of helping them build relationships between their practical knowledge and the specific teaching context.

Meanwhile, the peer coaching activities allowed the in-service teachers to learn from each other and to reflect on their own work. During the diagnosis and interactive activities, they benefited from others’ lesson plans, gave suggestions to each other, and reflected on their own plans. The knowledge structures of the in-service teachers could be promoted during the conflict and negotiation activities (Zhou, 2012). The social cognitive activities during peer coaching helped the in-service teachers go through an implicit-explicit-implicit transformation and iteration process which would benefit their understanding of the knowledge, develop their practical knowledge, and promote their instructional design and in-practice teaching abilities.

Using an empirical research method, this study explored the influence of the peer coaching-based personalized learning approach on in-service teachers’ learning participation, learning design skills, and in-practice teaching abilities. The results revealed that the peer coaching-based personalized learning approach had a positive influence on the three dimensions mentioned above, especially promoting the in-service teachers’ learning design skills and in-practice teaching abilities.

Although the proposed approach benefited teachers in this application, there are some limitations to this study. The number of participants was not enough, and the data used in analyzing the learning participation was not abundant. However, this study still provides a good reference for those who intend to conduct learning activities and studies related to the use of online peer coaching and personalized learning in in-service teachers’ professional development.

In the future, the assessment tools for precise diagnosis, online activities for peer coaching, and supporting tools for interaction and online learning can be improved by analyzing and constructing an accurate learner model and database. The more we know the teachers, the better we can promote the resources, activities, and supports for them. In addition, the proposed method can be adopted with mobile devices, on which in-service teachers can review their performance or give immediate feedback to others.

Acknowledgements

This research was funded by the Humanities and Social Science Research General Project of the Ministry of Education, China (Grant 15YJC880050), and the 2016 Comprehensive Discipline Construction Fund of the Faculty of Education, Beijing Normal University.

The authors would like to state that the subjects of this study were protected by hiding their personal information during the learning activity; moreover, participating in the experiment was voluntary and would not affect their grades. In addition, the subjects were allowed to withdraw from the experiment at any stage. The authors are pleased to provide the experimental data upon request and would like to state that there is no potential conflict of interest in this study.

References


Peer Assessment of Webpage Design: Behavioral Sequential Analysis Based on Eye Tracking Evidence

Ting-Chia Hsu*, Shao-Chen Chang* and Nan-Cen Liu
Department of Technology Application and Human Resource Development, National Taiwan Normal University

*Corresponding authors

ABSTRACT
This study employed an eye-tracking machine to record the process of peer assessment. Each web page was divided into several regions of interest (ROIs) based on the frame design and content. A total of 49 undergraduate students with a visual learning style participated in the experiment. This study investigated the peer assessment attitudes of the participants and found that they possessed highly positive attitudes towards and understanding of peer assessment. After comparing the results of the peer assessments and evaluation by experts, high consistency occurred when the design of the web page was concise; however, the consistency decreased when the web page content was too diverse. After comparing the peer assessment attitudes of the participants and their web page design scores, it was found that the web pages with concise designs attracted the visual-style students’ attention more, and that there was a significant negative correlation for those students who possessed a more negative attitude toward peer assessment. In addition, the study further analyzed the visual-style students’ serial behaviors in the peer assessment process for each web page design. After comparing the evidence of each student’s eye movements and his/her evaluation results, it was found that the students who gave higher or lower scores had different eye movements. For the website scored as having the best design, the fixations and behaviors of the assessors giving higher scores were highly consistent with those of the experts, implying that the few assessors giving lower scores were relatively poor at peer assessment. On the contrary, for the website which was scored as having the worst design, the fixations and behaviors of the assessors giving lower scores were highly consistent with those of the experts. Consequently, from the eye fixation hotspot evidence, when the students were more concentrated on the peer assessment, their evaluated results were closer to those of the two experts. Finally, the study found that the eye fixation hotspots were the same as the key points planned by the student designers of the website which scored the highest, which provided the student designers with additional important eye-tracking feedback from the peer assessment activities.

Keywords
Peer assessment, Eye tracking, Region of interest, Hotspots, Serial behavior

Introduction
The adoption of peer assessment has become a trend in the classroom (Boud, Cohen, & Sampson, 1999), and there is already much research indicating that peer assessment promotes the effective development of cognition and affect. However, the quality and fairness of the assessments by those students who lack experience and confidence when assessing their peers’ work have been questioned by many scholars (Ballantyne, Hughes, & Mylonas, 2002; White, 2009; Seng & Hill, 2014). Previous studies have also revealed that peer assessment is influenced by differences in the students’ culture, personality, and friendship (Johnston & Miles, 2004; Panadero, Romero, & Strijbos, 2013; McLeay & Wesson, 2014). Therefore, how to overcome the problem of participants’ individual differences in the peer assessment process has become an important issue.

Meanwhile, a growing number of studies have been exploring the impact of individual differences on peer review. Liu, Lin, and Yuan (2002) investigated the differences in the executive thinking styles of students in the peer assessment process. Learning style is one of the individual differences in digital learning studies. Solomon and Felder’s Learning style model is one of the famous models, and includes four facets (Felder & Silverman, 1988; Soloman & Felder, 2005), namely action/reflective, sensing/intuitive, visual/verbal, and sequential/global. The past learning style survey results or eye movement study results have shown that most students tend to adopt a visual learning style (Hsu, Hwang, & Chang, 2014; Hsu, Hwang, Chang, & Chang, 2013). However, one past study indicated that students with mixed learning styles constitute a higher proportion than single-style learners (Lujan & DiCarlo, 2006).

There is another important factor affecting peer assessment outcomes, namely students’ attitudes. There is a positive correlation between a positive attitude towards peer assessment and higher learning achievement. Therefore, in order to ensure the validity of peer assessment, teachers need to consider the students’ personal affect during the learning process (Lin, Liu, & Yuan, 2001). One past study also found that students usually have a positive attitude towards peer assessment activities. These positive attitudes ensure that the students are fair and responsible when completing peer assessment (Cheng & Warren, 1997). However, some students also worry...
about their ability and their responsibility when evaluating their peers’ work. To solve this problem, teachers must provide appropriate support to alleviate student stress (Cassidy, 2006).

In the last decade, many researchers have begun to analyze students’ learning behavior. Because of the different learning behaviors which affect students’ learning performance during the learning process (Hwang & Chen, 2016), learning behavior analytics has become an important issue in education (Hwang, Chu, & Yin, 2017). Learning behavior refers to the record of data related to students’ behaviors and interactions with peers during learning activities (Hwang, Hsu, Lai, & Hsueh, 2017). Hwang, Hung, Chen, and Liu (2014) emphasized that instructors could effectively improve their teaching methods and learning activities by analyzing learning logs or educational data. However, there are few or even no studies related to analyzing behaviors during peer assessment.

Therefore, this study used quantitative and qualitative evidence to explore the results of peer assessment by students with the same kind of learning style (i.e., visual learning style). When exploring the individual behavior of students with the same kind of learning style in the website reading process, the scientific evidence of the eye tracker can be more objective than just knowing the results of the peer assessment score. Moreover, the quantitative results of peer assessment can be used to compare the differences between the learning behaviors of students who give high and low evaluation scores. This study also investigated the correlation between the goal of the students’ attention and the goal (region) of the peer assessment process, and whether it was related to the students’ attitudes towards peer assessment. We could then count the eye-catching hotspots, which can provide another kind of important eye movement information feedback to the web page designers.

The research questions are as follows:
- Was the assessment of the student reviewers similar to the expert appraisal? What were the student reviewers’ attitudes towards peer assessment?
- What were the differences in the behavioral sequential analysis of eye movements during the peer assessment for the four websites with different designs?
- How did the behavioral patterns of the students giving peer assessment with low scores differ from those of the students giving peer assessment with high scores?

**Literature review**

**Eyes movement tracking**

The rapid development of eye tracking technology is based on the eye-mind hypothesis. Several researchers have found that one’s eye gaze is closely related to attention when dealing with visual messages (Just & Carpenter, 1976). Several scholars have proposed the premotor theory to explain the relationship between eye movement and attention. The main concept of the premotor theory is concentrated on the target, and uses eye movement to aim at the target to capture the visual behaviors. That is, attention is the pre-treatment before eye movement (Hoffman & Subramaniam, 1995).

At the same time, some scholars have referred to the Noticing Hypothesis to explain the relationship between attention and cognition. Schmidt (1990) considered that learners must be aware of the message input to ensure that learning is happening. It also means that learners must concentrate on learning to acquire knowledge when they are learning (Schmidt, 1990; Schmidt, 1993; Schmidt, 1995; Schmidt, 2001). In addition, attention can determine what information will be transferred from short-term into long-term memory during the cognitive process. That is, it is clear that attention is the key to learning (Egi, Fujii, & Tatsumi, 2002). Thus, attention and eye movement are closely related, and the development of eye tracking technology provides an additional reference for exploring previously unknown areas (Godfroid, Housen, & Boers, 2010).

The process of eye movement consists of two elements, the saccade and fixation. The saccade is the action of the eye moving, and fixation is the action of the eye stopping its movement and beginning to get the message. Fixation falls within the range of 100 to 500 milliseconds, depending on the target material. Eye movement is therefore constituted of a series of gazes and jumps (Rayner, 1998; Rayner, 2009). For eye tracking studies, regions of interest (ROIs) are often used by researchers to record visual behavior. ROIs will have definitions that are appropriate for different research questions. There are four common types of eye movement data, namely duration of the first fixation (DFI), latency of the first fixation (LFF), number of fixations (NOF), and total contact time (TCT). It is also possible to further measure the course of saccades (COS) in order to investigate...
students’ eye movement behavior. This is the most important eye movement data collected in this study (Hewig, Trippe, Hecht, Straube, & Miltner, 2008).

In past studies, some scholars adopted eye tracking technology to develop the adaptive digital learning (Adaptive E-Learning through Eye Tracking, AdELE) framework, using it to synchronize the recording of learning content and user information. Barrios et al. (2004) investigated users’ performance, knowledge level, and eye movement behavior to achieve an adaptive learning environment. In order to help the learners more, the eye tracking will record the user’s eye movement behavior instantly to provide the corresponding knowledge in different areas (Calvi, Porta, & Sacchi, 2008; Gütü et al., 2005; Pivec, Trummer, & Pripfl, 2006).

Tsai, Huang, Hou, Hsu, and Chiou (2016) utilized eye-tracking technology to explore the differences between high and low conceptual comprehension of university students’ visual behaviors and game flow in game-based learning. The results indicated that the students in the high-comprehension group showed efficient reading strategies for text and better metacognitive controls of visual attention during game play. Furthermore, Lindner, Eitel, Strobel, and Köller (2017) designed a science learning activity to record the eye movements of 62 schoolchildren solving multiple-choice (MC) questions. The results showed that the time the students spent fixating on the picture was compensated for by less time spent reading the corresponding text. In text-picture items, students also spent less time fixating on incorrect answer options.

Yang, Huang, and Tsai (2016) attempted to explore the effects of epistemic beliefs and the gender differences in epistemic beliefs in science on science-text reading. They administered the scientific epistemological beliefs (SEBs) questionnaire to the students and recorded the science-text reading process of the eye-tracking. The results demonstrated interactions between SEBs and gender, and the complicated SEBs were associated with higher cognitive attention to the reading of data-related information. Scholars have reviewed the eye-tracking studies and concluded seven learning-related categories employing the eye-tracking machine, namely patterns of information processing, effects of instructional strategies, re-examination of existing theories, individual differences, effects of learning strategies, patterns of decision making, and social effects (Lai et al., 2013).

From the above literature, we can see that eye tracking technology has been widely used in the field of e-learning. In addition to giving the students feedback with simple criteria items, peer assessment should pay attention to the differences between individuals. Hence, in this study, we used eye tracking technology to explore the differences between the learning styles and learning behavior of students.

**Peer assessment**

Peer assessment (PA) is a strategy for individuals to assess the work or learning outcomes of peers at the same level, helping students improve their cognitive ability (Topping, 1998). In the peer assessment process, all of the students not only act as the assessor, but also as the assessee (Van Zundert, Sluijsmans, & Van Merriënboer, 2010; Li, Liu, & Steckelberg, 2010). The characteristic of peer assessment is that it gives the students specific tasks and focuses on peer performance rather than on individual abilities. In this process, peers can be in the same or different classes, or be in different teams when executing the peer assessment activity (Lui & Andrade, 2014). Moreover, students’ understanding of their peers’ ideas during the learning process is also enhanced, and their weaknesses improved (Cho & Cho, 2011).

In the past decade, much research has reported the benefits of applying peer assessment to different skills training in school settings, such as communication skills (Lai, 2016; Wang, Liang, Liu, & Liu, 2016), writing skills (Cheng, Liang, & Tsai, 2015; Ashton & Davies, 2015; Tenório, Bittencourt, Isotani, Pedro, & Osipina, 2016; Russell, Van Horne, Ward, Bettsis, & Gikonyo, 2017), and higher-order thinking ability (Gielen & De Wever, 2015; Wang, Hou, & Wu, 2017). Peer assessment has been confirmed as being able to improve students’ learning performance, especially their writing skills (Cheng, Liang, & Tsai, 2015). In addition to improving their knowledge, students also respect the opinions of their peers and share their experience with others in the peer assessment activity (Tayem et al., 2015; Yu & Wu, 2013). Peer assessment will not only improve the learning outcome in the course, but will also increase social interaction and extracurricular discussion outside the class. Moreover, students’ high learning motivation will enhance their self-learning behavior (Xie, 2013). Peer assessment activities can also reduce the burden of teachers, provide opportunities for self-learning, and increase learning motivation and high-level thinking ability (Luo, Robinson, & Park, 2014).

Lai and Hwang (2015) proposed an interactive peer-assessment criteria development approach to help students make reflections while viewing their peers’ work. A total of 103 elementary students participated in their activity.
and were assigned to an experimental group and a control group. From the study results, it was found that the proposed approach significantly improved the students’ learning achievement and their learning motivation. It was also indicated that engaging students in assessment criteria development is an effective approach. Moreover, Wang, Hou, and Wu (2017) designed a 4-week collaborative learning activity involving four widely used instructional strategies, namely problem solving, peer assessment, role playing, and peer tutoring. The results showed that the students exhibited relatively more instances of the cognitive process of understanding when using the peer assessment and peer tutoring instructional strategies with Blogs. In other words, peer assessment can improve students’ ability to listen to suggestions and to discuss with peers.

Recently, Lai (2016) proposed an online video peer review system to support nursing students’ training in communication skills. A total of 50 Taiwanese nursing students attending a Psychiatric Care program participated in the study. In the learning activity, the students were asked to treat a simulated patient and record a YouTube video to have peers watch, rate, and give feedback on it. In the study, it was found that online peer assessment can be effective in nursing communication skills training, as their communication became more patient-centered.

Although many studies have reported the effectiveness of peer assessment in the last decade, several researchers have indicated that instructors should provide specific guidelines for how to decide a rating with subjective consciousness (Tenório, Bittencourt, Isotani, & Silva, 2016; Ng, 2016; Russell, Van Horne, Ward, Bettis, & Gikonyo, 2017). Therefore, the assessment criteria in the current study leave some flexibility for the students, such as allowing them to select “maybe” for certain items. Russell et al. (2017) investigated students’ evaluating processes and their perceptions of peer assessment when they were engaged in peer assessment using Calibrated Peer Review. They suggested that instructors may provide specific learning scaffolding for students of mid- or low-level ability.

Learning behavior analysis

According to Huang et al.’s (2014) definition of learning behavior analysis, students’ behaviors in learning activities can be recorded to analyze the factors affecting their learning performance. In recent years, much research has reported the benefits of learning behavior analysis for improving learning activities, such as in language courses (Chen & Yeh, 2017; Hwang, Hsu, Lai, & Hsueh, 2017), social networks (Darban & Polites, 2016; Buckley & Doyle, 2017), computer science (Su, Yang, Huang, & Tern, 2014), and science (Chiang, Yang, & Hwang, 2014).

For example, Hwang, Hsu, Lai, and Hsueh (2017) proposed a problem-based English listening game based on progressive sequential analysis to investigate the learning behavioral patterns of students with different levels of English anxiety. The experiment results indicated that high levels of English anxiety influenced students’ learning achievement and gaming behaviors. Moreover, Yin, Okubo, Shimada, Oi, Hirokawa, and Ogata (2015) proposed an e-book system which recorded the students’ learning logs of their daily academic life in the cloud. They found that the students’ learning behaviors were significantly related to their learning performance.

In addition, Chen and Yeh (2017) developed an online test to help students learn Academic English, and investigated the effects of the students’ cognitive styles in the English learning context. The results of the quantitative measurement indicated that cognitive styles have significant influences on students’ learning patterns in the context of Academic English. However, there is a lack of studies investigating how learning styles affect students’ reactions in the context of peer assessment. To address this issue, we designed an eye tracking activity to investigate the sequential behaviors of students with different learning styles in a peer assessment activity.

Learning styles

Learning styles are defined as the individual learning habits, methods, and attitudes of a learner, and are influenced by the learner’s interactive skills and the learning environment (Dunn & Dunn, 1987). The study of learning style is derived from the cognitive style of psychology; the most important aspect of cognitive style is learning style. Therefore, the measurement of learning style should focus on the various learning activities that determine the individual’s cognitive style (Sternberg & Grigorenko, 1997). Because of the rapid development of learning style research in recent years, the term learning style has become widely used.
Many scholars have begun to advocate the learning style model, because it can improve the plight of higher education teaching by helping teachers engage in more profound thinking and face the diversity of students (Claxton & Murrell, 1987). There are other scholars who also believe that in higher education it should not be assumed that all students can use the same way of learning. Teachers should be responsible for increasing their adaptive teaching methods to allow different kinds of students to achieve more effective learning (Hawk & Shah, 2007). That is, when teachers are designing courses and preparing teaching materials, they need to consider how to help students develop their interpersonal relationships, leadership, and communication skills.

Felder and Silverman (1988) proposed a learning style theory which is compatible with different learning styles and teaching methods. Based on that theory, Soloman and Felder (2005) further proposed the Index of Learning Styles (ILS), which includes the four dimensions of active/reflective, sensing/intuitive, visual/verbal, and sequential/global learning. In addition, some scholars have attempted to study the psychological model (such as human behavior and learning style) in information technology. The results showed that the Felder-Silverman learning style model has the advantages of simplicity and comprehensiveness, and is well suited for use in digital learning (Fatahi, Moradi, & Kashani-Vahid, 2016). For example, some scholars have proposed a model that examines the mediating processes in the relationship between learning style and e-learning performance. The results showed that the sensory/intuitive dimension learning style will indirectly affect learning performance through the regulation of online participation, while other types of learning styles do not have any effect (Huang, Lin, & Huang, 2012).

Therefore, this study attempted to combine learning style and eye tracking to investigate students’ sequential behaviors in a peer assessment activity. Through the learning style and eye tracking sequence analysis of the records, it was hoped to understand the students’ learning situation, in order to provide relevant research results and recommendations for future researchers.

**Method**

**Participants**

The study participants were undergraduates from a university in the north of Taiwan, and were approximately 20 to 21 years old. There were 80 students who had taken the e-learning course before this activity. After deducting the missing values, 57 participants with complete data remained. According to the learning style scale (Index of Learning Styles) analysis, 49 of the participants were visual learning style students (21 female and 28 male). None of these 49 students had used an eye-tracking machine before.

**Measuring tools**

The questionnaire of learning styles was adopted from the measure developed by Soloman and Felder (2005). It consisted of 44 items for evaluating learners’ dimensions of learning style including active/reflective, sensing/intuitive, visual/verbal, and sequential/global. Each question has two opposite answers, which represent different individual learning styles. In this study, we only used the 11 visual/verbal dimension questions to measure whether or not the students adopted a visual learning style (Soloman & Felder, 2005).

<table>
<thead>
<tr>
<th>Table 1. The Kelly repertory grid designed by the website experts</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Different website features</strong></td>
</tr>
<tr>
<td>The framework of the upper and lower pages.</td>
</tr>
<tr>
<td>The framework of the left and right pages.</td>
</tr>
<tr>
<td>There is no frame page.</td>
</tr>
<tr>
<td>The website interaction (Such as bulletin boards, contact boxes, information management, business-related websites and resource links, etc.)</td>
</tr>
<tr>
<td>The color coordination of the web pages.</td>
</tr>
<tr>
<td>The layout is simple.</td>
</tr>
<tr>
<td>The text is interesting.</td>
</tr>
</tbody>
</table>

*Note. 5 = Strongly agree, 1 = Strongly disagree.*

In this study, four of 10 groups of students’ websites designed in the web design course were selected, each representing one of four different themes. The first website was designed to promote pearl beads and sell brand-
related products. The second website was an introduction to the school dance club organization and activities. Another website was an introduction to the famous night markets and restaurants near the campus. The last one is a website designed to sell second-hand clothing. Each website consisted of 10 web pages. The four websites were selected for the peer assessment activity, and the students scored them in accordance with the rating scale designed by the expert, as shown in Table 1.

Table 2. The rating scale of web peer assessment

<table>
<thead>
<tr>
<th>Score item</th>
<th>Please select your rating of the students’ website</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>YES (Positive)</td>
</tr>
<tr>
<td></td>
<td>GOOD</td>
</tr>
<tr>
<td></td>
<td>BAD</td>
</tr>
<tr>
<td></td>
<td>VERY BAD</td>
</tr>
<tr>
<td>The font size of the text.</td>
<td></td>
</tr>
<tr>
<td>The placement of images.</td>
<td></td>
</tr>
<tr>
<td>The hyperlink location design and arrangement.</td>
<td></td>
</tr>
<tr>
<td>The layout will change when the user changes the page.</td>
<td></td>
</tr>
<tr>
<td>The logo or trademark is obvious.</td>
<td></td>
</tr>
<tr>
<td>The background of the website is appropriate.</td>
<td></td>
</tr>
<tr>
<td>It is easy to find the important information on the web page.</td>
<td></td>
</tr>
<tr>
<td>Do you feel that you were a concentrated reviewer during the peer assessment?</td>
<td></td>
</tr>
</tbody>
</table>

The peer assessment scale was developed by two experienced professors who have designed e-learning digital content for several years, as shown in Table 2. The Cronbach’s α value of internal consistency reliability was .846, showing that the peer assessment was highly reliable. After we finished the design of the peer assessment scale, we also scored the four web pages. The Kappa consistency score of reliability was 0.903, showing that there was a high degree of consistency between the two experts.

In this study, we modified the college students’ peer assessment scale developed by Wen and Tsai (2006). The purpose was to investigate students’ attitudes and acceptance of the peer assessment activities. The Cronbach’s alpha value of the analysis result was .75, showing good reliability in internal consistency.

Research design

Before the test, the participants were divided into two groups according to their different learning styles and the assessment sequence to complete the peer assessment of the four websites. In order to avoid the hangover effect of peer assessment and tired eyes, we used a counterbalanced design, as shown in Table 3.

Table 3. Counterbalanced design

<table>
<thead>
<tr>
<th>Random</th>
<th>1st time of eye tracking</th>
<th>2nd time of eye tracking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>Website A → Website B</td>
<td>Website C → Website D</td>
</tr>
<tr>
<td>Group 2</td>
<td>Website C → Website D</td>
<td>Website A → Website B</td>
</tr>
</tbody>
</table>

Figure 1 shows the experimental design of this study. Before the learning activity, the two groups of students were instructed in digital multimedia design to learn its basic knowledge, which is a part of the learning course. Before the experiment, we asked the students to complete the learning style questionnaire. We then explained the eye tracking instructions and the eye tracking correction. During the learning process, the eye tracking machine recorded each user’s process of peer assessment as they completed the peer assessment scale for the web page. After that, the students rated another website, and each website was rated two times until the students completed all of the ratings. After the learning activity, the students completed the questionnaire to measure their learning attitudes towards peer assessment.
Figure 1. The process of the experimental activity

Research instruments

The research equipment used was the Eye NTNU-180 (a cheap multi-screen eye tracker) developed by the National Taiwan Normal University, Department of Electrical Engineering research team. This eye tracking instrument uses infrared LED detection of eye gaze focus on objects, and has a sampling frequency of between 30 and 180 Hz within 60 cm. The error rate is less than 0.3 degrees, and there is a 9-point correction technique to improve accuracy. An example is shown in Figure 2.

Figure 2. The experimental equipment and environment

As each web page has a frame, we categorized the text and the image area of the page frame as an ROI. We analyzed the data of the visual students in the peer assessment process and summed the scores excluding the personal attentiveness score. According to Kelley’s (1939) definition, the top 27% of the students were classified as the high evaluation group, and the bottom 27% were classified as the low evaluation group (Kelley, 1939). The number of students reviewing the four websites is shown in Table 4.
Results

Quantitative results of peer assessment of the webpage designs

The students showed that they had concentrated on the process of peer assessment for each website. Based on the 4-point Likert scale ranging from 1 to 4, the self-report inventory showed that the reviewers’ average concentration degrees for websites A, B, C, and D were 3.20, 3.04, 3.14, and 3.22 respectively during the peer assessment. In other words, the students presented high confidence in their concentration on the process of peer assessment.

The performance of the four websites was also ranked by the experts. The ranking from better to worse was A, D, B, C. After comparing the assessments of the experts and the student-reviewers, it was found that the consistency between their scoring decreased when the webpage design was worse. The similarity degree between the evaluation results of the student reviewers and the expert reviewers was more than 85% for websites A and D. Accordingly, the students’ evaluation was closer to that of the experts when the webpage had a better design.

On the contrary, the evaluation of the student reviewers had about 60% similarity with the assessment of the expert reviewers for websites B and C, as shown in Table 5. Therefore, the grades given by the students were higher than those given by the professors when the webpage was not well designed. The experts gave stringent evaluations and made clear judgements, while the student reviewers seemed to be concerned about their peers’ feelings. The students tended to give their peers higher grades when their proficiency was not professional enough to judge the poor design of the website, or if there was too much diversity in the design presented in one website. For example, the background of website B was so bright that the information displayed was relatively neglected.

Table 5. The consistency percentage of the experts’ and student reviewers’ assessments per evaluated category.

<table>
<thead>
<tr>
<th>Material Website</th>
<th>Text</th>
<th>Picture</th>
<th>Link</th>
<th>Change page</th>
<th>LOGO</th>
<th>Background</th>
<th>Information</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>75.51%</td>
<td>97.96%</td>
<td>97.96%</td>
<td>87.76%</td>
<td>87.76%</td>
<td>93.88%</td>
<td>89.80%</td>
<td>90.09%</td>
</tr>
<tr>
<td>B</td>
<td>59.18%</td>
<td>51.02%</td>
<td>71.43%</td>
<td>57.14%</td>
<td>36.73%</td>
<td>87.76%</td>
<td>28.57%</td>
<td>55.98%</td>
</tr>
<tr>
<td>C</td>
<td>73.47%</td>
<td>51.02%</td>
<td>36.73%</td>
<td>44.90%</td>
<td>46.94%</td>
<td>53.06%</td>
<td>51.02%</td>
<td>51.02%</td>
</tr>
<tr>
<td>D</td>
<td>79.59%</td>
<td>97.96%</td>
<td>93.88%</td>
<td>71.43%</td>
<td>91.84%</td>
<td>75.51%</td>
<td>85.71%</td>
<td>85.13%</td>
</tr>
</tbody>
</table>

It was found that 27 students achieved 70%-79% similarity with the experts, which constituted the majority of the students. Secondly, 14 students had 60% - 69% similarity in comparison with the assessment of the experts. However, there was no correlation between the peer assessment attitudes of the students and the similarity degree of the students’ peer assessment and the experts’ evaluation, as shown in Table 6.

Table 6. The distribution of different consistency levels

<table>
<thead>
<tr>
<th>Proportion attitude</th>
<th>50%-59%</th>
<th>60%-69%</th>
<th>70%-79%</th>
<th>80%-89%</th>
<th>90%-100%</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>4</td>
<td>14</td>
<td>27</td>
<td>3</td>
<td>1</td>
<td>49</td>
<td>49</td>
</tr>
<tr>
<td>Positive attitude</td>
<td>3.72</td>
<td>3.53</td>
<td>3.16</td>
<td>3.67</td>
<td>2.25</td>
<td>3.32</td>
<td>0.56</td>
</tr>
<tr>
<td>Understanding-and-action</td>
<td>3.92</td>
<td>3.98</td>
<td>3.94</td>
<td>3.89</td>
<td>4.00</td>
<td>3.95</td>
<td>0.45</td>
</tr>
<tr>
<td>Negative attitude</td>
<td>2.31</td>
<td>2.84</td>
<td>2.42</td>
<td>3.17</td>
<td>2.00</td>
<td>2.57</td>
<td>0.71</td>
</tr>
</tbody>
</table>

In terms of the peer assessment attitudes, this study utilized a one-sample t test to examine whether each dimension of peer assessment was significantly larger or smaller than the middle of the 5-point Likert scale (i.e., 3). The results showed that 49 students had significantly positive attitudes towards peer assessment with mean = 3.32 and $SD = 0.56$ ($t = 3.988^\text{***}$, $p < .001$), and showed the attitude of understanding-and-action for peer assessment with mean = 3.95 and $SD = 0.45$ ($t = 14.446^\text{***}$, $p < .001$). There was negative significance for negative attitudes towards peer assessment with mean = 2.57 and $SD = 0.71$ ($t = -4.252^\text{**}$, $p < .001$). Accordingly, the students presented high understanding and positive attitudes and showed lower negative attitudes towards the peer assessment activities.
Behavioral sequential analysis of eye movements

In this study, we evaluated the students’ saccade path (i.e., course of saccade, COS) among the regions of interest (ROIs) such as the text paragraph, the picture area, the menu area, the logo, or the region outside of the webpage information area. A total of 129,020 eye-tracking records were collected, with 27,224 for website A, 28,773 for website B, 36,015 for website C, and 37,008 for website D. Each webpage was divided into a maximum of five regions. After using the time serial sequential analysis method to find the significant paths \((z > 1.96, p < .05)\) of eye movement from one region to another, each page of every website was evaluated. The arrow is thicker when the \(z\) scores between the two regions are larger. The \(z\) scores are larger when the students saccade from one region to another more frequently. For example, the behavioral sequential patterns for the first page of website A are shown in Table 7. Overall, from the results of the behavioral analysis of website A, the students put most of their focus on the three main areas. They read the logo first, then read the menu, and finally read the following information. Consistent behaviors were found for seven of the 10 webpages of website A.

Table 7. Behavioral sequential analysis of eye movements

<table>
<thead>
<tr>
<th>Website</th>
<th>Behavioral serial patterns</th>
<th>Webpage template</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>T11 → T12 → T13 → E1 → T13</td>
<td><img src="image" alt="Webpage A" /></td>
</tr>
<tr>
<td>B</td>
<td>T31 → T32 → T33 → E1 → T11</td>
<td><img src="image" alt="Webpage B" /></td>
</tr>
<tr>
<td>C</td>
<td>P11 → T12 → T13 → E1 → P11</td>
<td><img src="image" alt="Webpage C" /></td>
</tr>
</tbody>
</table>
As for website B, there were too many COS (course of saccade) behaviors. The students spent more effort moving their eyes among the different frames. From the overall saccade behaviors of website B, too many frame pages combined in one web page did not make for a good design because the students were not able to ignore unnecessary information (e.g., the fancy background) and focus on the particular point that the designer wanted them to pay attention to.

In terms of website C, the students mainly moved their eyes between the title and the explanations. Therefore, when the two pieces of information were correlated, the students got used to matching the corresponding areas, resulting in the eye behaviors moving back and forth. One behavior was the same for websites B and C. The students scanned around the space in the web page. They may have been searching for more information.

Overall, in website D, a logo and menu which were put in a fixed place inferred that they had a more conspicuous design which was able to attract attention. The explanations had best not be too far from the pictures so that the students could easily scan between the pictures (e.g., the clothes pictures) and their corresponding explanations.

Eye fixations and saccades of different assessors

It was found that the regions which the reviewers giving higher scores gazed at were different from the regions which the reviewers giving lower scores fixated on. What they saw became a reference or reason for why they gave their peers’ website high or low scores. This study explored whether the students giving high or low scores evaluated each website validly. The student designers provided the key regions of their design plan in each web page when they published the website. The key regions referred to the key points on the page which were expected to attract the viewers’ attention. In this study, we compared the fixation heat map with the key points of the design plan.

For example, on the first page of website A, the students giving high scores frequently moved their eyes from the logo to the outside space area. From the saccade pattern, it was found that the student reviewers gave evaluation with lower scores when they gazed at the text at the bottom and looked around the space area, as shown in Table 8. The same behaviors of the student reviewers doing peer assessment and giving high and low scores were removed. Therefore, the different saccade behaviors of the student reviewers doing peer assessment with high and low scores were found. In addition, we also examined the eye hotspots of the student reviewers doing peer assessment with high and low scores. For example, in Table 8, it could be inferred that the ROI of T11 on website A may be the reason why points were deducted for the webpage, as other ROIs had similar percentages of fixation from the reviewers, but T11 had almost double the number of reviewers giving it lower scores. After interviewing the reviewers, they said that the logo was a little too small so they gazed at it more frequently at first.

On the contrary, when the web page design was poor, the careful reviewers gazed at different areas to find out where its key point was. For example, website B had a relatively poor web page design because its background could not clearly show the information content. Therefore, the students giving lower scores for website B had more patterns of saccades among the ROIs, as shown in Table 9. A comparison of the hotspots of the reviewers giving peer assessment with high and low scores showed that some students gave lower scores because they could not gaze at the numerous words for a long time because the design distracted them from the content. The eye-tracking provides triangulated evidence which confirms the scoring process.
Table 8. The different behaviors of the reviewers who gave peer assessments with high and low scores for an example page of website A

<table>
<thead>
<tr>
<th>Different behaviors</th>
<th>An example of ROIs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixation partition of the reviewers who gave peer assessment with high scores</td>
<td>Fixation partition of the reviewers who gave peer assessment with low scores</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>T11</th>
<th>T12</th>
<th>T13</th>
<th>E1</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>7</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>30%</td>
<td>70%</td>
<td>70%</td>
<td>60%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>T11</th>
<th>T12</th>
<th>T13</th>
<th>E1</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>10</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>57%</td>
<td>71%</td>
<td>71%</td>
<td>43%</td>
</tr>
</tbody>
</table>

Note. Red solid lines refer to the unique behavioral patterns of the students giving peer assessment with high scores. Blue dotted lines refer to the unique behavioral patterns of the students giving peer assessment with low scores.

In sum, in terms of website A, seven of the 10 webpages had high scores with frequent and effective fixations on key points in the page during the peer assessment. From those behavior patterns, it could be concluded that the students giving high scores conducted an effective assessment of website B. With regard to the evaluation results given by the experts, the overall rating of website A was also high. The triangulated evidence confirmed that website A should be ranked top among the four websites.

However, as for website B, only three of 10 webpages had high scores with frequent fixations on key points at the same time. Most students assessed website B with low scores, and those students concentrated on the design key points. In other words, the students giving low scores performed effective assessment, rather than those giving high scores. The overall ranking of website B was not so good according to the evaluation results of the experts. As for website C, it was worse, as seven webpages were given low scores with effective eye movement records. Website D had five web pages with high scores and the other five had low scores. The overall ranking of the four websites from high to low scores in the peer assessment was A, D, B, and C, which matched the ranking given by the experts.

On the other hand, the number of student reviewers gazing at the same ROIs as the experts was defined as concentration. This study selected the top 27% concentrated student reviewers (i.e., \( N = 48 \times 0.27 = 13 \)), and the lowest 27% concentrated student reviewers (i.e., \( N = 48 \times 0.27 = 13 \)) from all participants based on the fixation analysis. The results are displayed in Table 10. The experts assessed websites A and D as high scoring, while they appraised websites B and C as low scoring. The study explored how the high- and low-concentrated student reviewers scored the four websites. The evidence of the eye-tracking from the peer assessment process showed that website A had the best design of the four websites. The average of the scores given by the experts was not only the highest, but also the student reviewers all gave it high scores regardless of being low- or high-
concentrated. Website D had a similar situation. Conversely, websites B and C gained low scores no matter whether they were evaluated by high- or low- concentrated student reviewers.

Table 9. The different behaviors of the reviewers who gave peer assessments with high and low scores for an example page of website B

<table>
<thead>
<tr>
<th>Different behaviors</th>
<th>An example of ROIs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixation partition of the reviewers who gave peer assessment with high scores</td>
<td>Fixation partition of the reviewers who gave peer assessment with low scores</td>
</tr>
<tr>
<td>T31</td>
<td>T32</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>38%</td>
<td>46%</td>
</tr>
</tbody>
</table>

Note. Red solid lines refer to the unique behavioral patterns of the students giving peer assessment with high scores. Blue dotted lines refer to the unique behavioral patterns of the students giving peer assessment with low scores.

Table 10. The cross-match between the heat maps and scores given by the experts and peers

<table>
<thead>
<tr>
<th>Website</th>
<th>Results of expert assessment</th>
<th>Results of peer assessment</th>
<th>High concentration</th>
<th>Low concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>High scores</td>
<td>High scores</td>
<td>13</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>Low scores</td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>Low scores</td>
<td>High scores</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>Low scores</td>
<td></td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>C</td>
<td>Low scores</td>
<td>High scores</td>
<td>1</td>
<td>92.31%</td>
</tr>
<tr>
<td></td>
<td>Low scores</td>
<td></td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>D</td>
<td>High scores</td>
<td>High scores</td>
<td>12</td>
<td>92.31%</td>
</tr>
<tr>
<td></td>
<td>Low scores</td>
<td></td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Consequently, the low- and high-concentrated student reviewers all performed effective assessment because their evaluation achieved more than 90% similarity with the expert appraisal. The reason why the high-concentrated student reviewers were not able to achieve 100% effective assessment is inferred as follows. It could be said that high concentration is not the only necessary condition for highly effective peer review because the highly concentrated students may have focused on the areas which were not the key points of the page. When the students concentrated on the area which was not the main information place the designers wanted to deliver, they would not absorb the full message, meaning that some of their evaluation could not be effective or valid. Therefore, the heat maps alone could not ensure that the students’ assessments were valid. The heat maps had to be triangulated and examined along with the quantitative data including the students’ attitudes towards peer
assessment, the scores given by each student reviewer for every website, and the saccade patterns of the student reviewers.

Discussion

From the quantitative results, it was found that the consistency rate between the evaluation performed by the experts and that of the peers achieved more than 70%. This conforms to a previous study (Falchikov & Goldfinch, 2000) which indicated that providing student reviewers with comprehensive evaluation criteria or rubrics for assessing products could result in evaluation results like those of experts. A previous study has shown that understanding the assessment rubrics well could ensure that students complete the peer review fairly (Cheng & Warren, 1997). The investigation results of peer assessment attitude in the current study showed that the students possessed positive attitudes towards and understanding of peer assessment in this study. However, it should be noted that this study did not introduce the comments given by the students and experts. Future studies are encouraged to code such comments and to compare them with the eye movement clues.

After confirming that most students’ assessment capabilities were close to those of the experts with the help of clear rubrics, it was then found that the assessment results were also affected by the diversity of the materials presented in the websites from the behavioral sequential analysis of the eye movements. In particular, poor web interface design caused a lower degree of consensus between the assessment of the experts and students. This study provides additional evidence and triangular validation for a previous study on the eye tracking technique (Iqbal & Mahmood, 2008), revealing that the assessment of teachers would be different from that of students when the students and teachers evaluate a disappointing creation (Iqbal & Mahmood, 2008).

The current study found that the grades given by the students tended to be higher than those given by the experts when the students were confronted with a design they were not familiar with. Another study indicated that the students gave higher scores in comparison with the teachers when the students faced an unfamiliar scoring item or target (Lindblom-ylänne et al., 2006). Based on this finding, it is suggested that teachers should avoid only exhibiting well-designed websites, but should increase the diversity of examples so that students could not only learn the advantages of well-designed webpages but could also think about how to improve poor designs. The appraisal capability of the students could be enhanced by such training.

In this study, it was also concluded what constitutes a better design for a web page. Website A was more concise and clearer than website D. From the heat map, it was found that the viewers were more concentrated when they evaluated Website A. Website A did not cause too much multimedia cognitive loading due to its clear design (Mayer & Moreno, 2003). A clearer design will reduce student reviewers’ negative attitude toward peer assessment and promote their concentration. In addition, the design of the logo and menu was consistent on each page of Website A. Most of the students looked at the logo and menu before looking at the content. The design was close to the experience of the students with a visual learning style (Axelsson, 2012), and similar to another reading study which pointed out that the students had movement of the eye between the picture and its explanation when the picture and text were related (Mason et al., 2015).

It is also concluded what constitutes a poor design for a web page. In Websites B and C, it was not easy for the assessors to distinguish the background and words on most of the pages. Scholars have mentioned that web design would be evaluated poorly when the visibility of important content is unclear (Fleming & Koman, 1998) as the student reviewers would be easily distracted from the focal point. In order to check whether the designers’ intentions are well presented on their web pages, the designers should point out the positions of the key points on each page. After comparing the key points with the viewers’ fixations, the reviewers could be sure that they had the right focus when observing the website and that the designers did successfully deliver their subject. The design was unsuccessful when the student reviewers could not focus on the information which the designers mainly wanted to deliver. Finally, the students not only had to concentrate on the web page but also had to focus on the right place (Ariasi & Mason, 2011).

Conclusions

This study incorporated the evidence of eye movements with peer assessment and expert assessment to triangulate the validation in order to confirm whether the students who gave high scores or those who gave low scores conducted an effective peer assessment for a particular website, and to explain which webpage design attracted the viewers’ attention to the right regions. We employed the eye movement technique to open the black
box of the peer review process. The analysis provides more scientific evidence to better understand the peer assessment process so that the reliability and validity of peer assessment could be enhanced.

The results indicated that the students giving higher and lower scores had different eye movements. When the higher scores given by the student reviewers were highly consistent with the evaluation of the experts, those who gave lower scores were relatively poor at peer assessment. Based on the eye fixation hotspot evidence, when the students were more concentrated on the peer assessment, their evaluated results were closer to those of the two experts. The eye fixation hotspots were the same as the key points planned by the student designers of the website which scored the highest. In other words, the eye-tracking behavioral records of peer assessment provided the students with additional important feedback for designing a website.

Finally, it should be pointed out that the major research limitation of this study is the limited number of times that peer assessment was performed. Future studies are encouraged to conduct more rounds of peer assessment so that the students have opportunities to repair their designs. The visual learning style was the control item in this study, so future studies could consider making use of different learning styles as a dependent variable.

Acknowledgements

This study is supported in part by the Ministry of Science and Technology in the Republic of China under contract numbers MOST 105-2628-S-003-002-MY3. This work is also partly supported by the “Aim for the Top University Project” of the National Taiwan Normal University and the Ministry of Education, Taiwan, R.O.C. The authors would like to thank Prof. Hong-Fa Ho who provided the eye-tracking equipment and technical support.

References


Effects of a Progressive Prompting-based Educational Game on Second Graders’ Mathematics Learning Performance and Behavioral Patterns

Kai-Hsiang Yang¹, Hui-Chun Chu²* and Li-Yu Chiang¹

¹Department of Mathematics and Information Education, National Taipei University of Education, Taipei, Taiwan // ²Department of Computer Science and Information Management, Soochow University, Taipei, Taiwan

// khyang.sinica@gmail.com // carolhcchu@gmail.com // liyu812@hotmail.com

*Corresponding author

ABSTRACT

Game-based learning (GBL) has been proven to be an attractive learning model by many studies; however, scholars have pointed out that the effectiveness of game-based learning could be limited if proper learning strategies are not incorporated. Prompting is a strategy that plays the important role of providing hints and guidance in interactive learning environments. Therefore, this study proposes a game-based learning approach with a progressive prompting strategy, using different levels of hints to guide students to complete tasks and achieve learning goals. Quasi-experimental research was employed in this study using two groups of students. The experimental group learned with the proposed approach, while the control group was allocated a conventional game-based learning strategy. The findings of the study show that the proposed approach significantly improved the second graders’ mathematics learning achievement. From the analysis of the experimental students’ learning behaviors, it was reported that the experimental group students could generate the answer after two progressive prompts. Thus, we could see that the proposed approach could enhance the learning achievement of the experimental group students by correctly guiding them to answer questions, step up their thinking, and understand the learning content in the learning process.

Keywords

Progressive prompting, Game-based learning, Learning behavior, Mathematics course

Introduction

With the advancements in information technology, the conventional teaching model has changed. In recent years, the combination of games and learning has gradually become a popular research topic of game-learning; many studies have indicated that game-based learning could enhance learners’ interest and motivation in learning (Gee, 2003; Ebner & Holzinger, 2007; Dickey, 2011; Moreno, 2012). Based on the game fantasy, curiosity, and sense of control for learners, their learning motivation could be stimulated, and thus their learning performance could be enhanced (van Eck, 2006). Game-based learning has been recognized as a potential learning model; however, previous studies have indicated that it may not be beneficial for learners without appropriate learning strategies in the process (Charsky & Ressler, 2011; Hwang, Sung, Hung, Huang, & Tsai, 2012; Wouters & van Oostendorp, 2013). The research findings of Mitchell and Savill-Smith in 2004, for instance, indicated that the learners’ attention to learning may be distracted if the purposes of the game and the learning are not identical in the game-based learning context. Thus, it has become very important to provide immediate guidance and hints to learners when game-based learning is conducted.

A well-designed game-based learning context could stimulate learners to challenge themselves by performing game tasks and solving learning problems repeatedly through immediate prompts before gaining subject knowledge. Moreover, learners may learn relevant knowledge and concepts that they were previously unfamiliar with in the problem-solving process (Kirriemuir, 2002). The progressive prompting strategy could help learners to solve problems based on various levels of hints. On the contrary, if learners could lose their learning motivation without the help of instant hints or guidance, it would eventually affect their learning achievement (Gibbs & Hrabshaw, 1989). Miller and Mercer (1993) proposed a concrete-semiconcrete-abstract (CSA) approach as a progressive prompting strategy for nine students with mathematics disabilities. Their research findings showed that the CSA is not only beneficial for students in terms of learning mathematics, but also helps with the learning retention.

Recently, many researchers have tried to design and implement feedback or guiding mechanisms, and most of the mechanisms have been confirmed as being effective in educational settings. For example, Chu et al. (2010) adopted a two-tier test to diagnose misconceptions, and provided adaptive feedback according to the diagnostic results to replace the misconceptions. Wang (2014) developed a graduated prompting online test system for providing the assessment feedback with three-step text prompts to assist learners’ learning. Chen, Hwang, and Tsai (2014) proposed a progressive prompt-based context-aware learning approach and effectively improved the students’ learning performance. They also found that the proposed approach encouraged the students to put more
effort into examining the contextual information and interpreting the learning content. Therefore, it is important to provide adequate prompts to assist students’ learning in the personal learning process.

Among various applications of technology-enhanced learning, Mathematics is an important subject due to its high correlation with our daily life. Imaging math situations will help learners concretize the content, and then they should be able to learn math more easily. However, as far as we know, there are few studies on feedback content presented in an image-based format. Accordingly, in this study, we developed an educational game with a progressive prompting mechanism, by providing two prompts which progressively moved from abstract (text description) to concrete (image description) for learners to construct the correct concepts, comfortably learn in the game-based learning environment, and finally consolidate their mathematical knowledge foundation. To evaluate its effectiveness, the following research questions were investigated to evaluate the performance of the proposed approach from various aspects:

- Did the students who used the game-based learning approach with the progressive prompting strategy show better learning achievements than those who used the conventional game-based learning approach?
- Are there differences between the learning achievements of students with different levels of mathematics self-efficacy in the experimental group?
- Are there differences between the flow experiences of the students who learned with the different learning modes (the game-based learning approach with the progressive prompting strategy and the conventional game-based learning approach)?
- Did the students who learned with the game-based learning approach with and without the progressive prompting mechanism show different learning behaviors?

**Literature review**

**Game-based learning**

Game-based learning (GBL) refers to a computer game environment in which learning and interactive entertainment are combined to create students’ learning fun and learning achievement based on learning theories (Prensky, 2007). In recent years, many studies have focused on GBL because games could make students become immersed in the game situations characterized by challenge and fun, thus improving the students’ motivation and helping them acquire subject knowledge. For example, Hwang, Chiu, and Chen (2015) designed a game-based learning context with “Saving Island” and “Investment Island” to help sixth graders learn financial concepts in a social studies course. Chu, Yang, and Chen (2015) developed an educational computer game for the unit of “Siege of Fort Zeelandia by Zheng Cheng-Gong” in an elementary school history course. They found that the approach improved the students’ learning achievement. Wang and Chen (2010) indicated that challengeable game tasks would bring high flow experience and high interest for learners participating in the games so that their learning achievement could be enhanced. Chang, Wu, Weng and Sung (2012) also reported that GBL could promote learners’ high learning motivation, more flow experience, and better learning outcomes in comparison with conventional approaches. Barzilai and Blau (2014) indicated that learning and enjoyment would lead learners to complete tasks successfully; moreover, both learning and gaming experience show high interaction and correlation in GBL.

To sum up, learners can gain enjoyment, self-confidence, and satisfaction from educational games if they are given challenging tasks that are equal to their skills and knowledge in GBL. Learning attitude and learning achievement can be improved when learners’ flow experience occurs in GBL. However, several studies have reported that although students’ learning motivation could be raised in GBL, no significant improvement was found in their learning achievement without incorporating proper learning strategies into the gaming process (Hwang & Chang, 2016; Charsky & Ressler, 2011). Thus, many studies have emphasized that it is important to embed teaching strategies or learning theories into GBL (Charsky & Ressler; Hwang, Yang, & Wang, 2013; Hsiao, Chang, Lin, & Hu, 2014). Hwang and Wang (2016) investigated students’ performance of learning English vocabulary with different guiding strategies in a game to serve as guidance for learners. The experimental results showed that the students using the game with the cloze guiding strategy had significantly better learning achievement than those learning with the multiple-choice guiding strategy. Thus, the current study proposes a game-based learning approach with different types of progressive prompting strategies. We hope the proposed approach could help learners complete game tasks in GBL.
Progressive prompting strategy

Progressive prompting is assistance progressively provided to a learner by teachers in order to increase the probability of correct responses. Teachers guiding students with scaffolding to solve problems and tasks is the most important principle of prompting. That is, prompting aims to allow learners to think how to solve problems and to challenge themselves to exceed their current ability. Wang, Huang, and Hwang (2016) indicated that if students cannot understand their own learning weaknesses and blindness, it is difficult for them to improve their learning situation. The role of the teacher is to guide students to solve learning tasks when they lack the necessary knowledge. As long as the learning tasks are completed, their performance could be improved (Mayer, 2004). With the advancements in information technology, learners are able to be given appropriate prompts by learning systems as soon as they encounter problems in learning activities. Therefore, subject course designers or teachers have to consider when and how to provide immediate and appropriate hints to help students think about and improve their problem-solving ability in the learning process (Chen & Choi, 2010). Chu and Chang (2014) developed an educational computer game system with a two-tier test mechanism to guide fifth graders to learn the bird identification unit of a natural science course. The experimental results showed that the proposed approach not only significantly promoted the students’ learning motivation, but that they were also able to correctly identify the learned knowledge of birds. In the researchers’ opinion, the advantage of the two-tier test is that it helps the learners revise their misconceptions or alternative conceptions. On the other hand, the progressive prompting strategy’s advantage is that it strengthens internalization of the concepts.

Many studies related to prompting strategies have been performed. For example, Gerber, Semmel, and Semmel (1994) developed a computer-based dynamic assessment of multi-digit multiplication with progressive prompting functions for secondary students. They found that most of the test takers who were able to solve the mathematics questions successfully needed more prompts as the difficulty of the test items increased. Chen, Hwang, and Tsai (2014) developed a progressive prompt-based context-aware learning approach with three-stage prompts. Their results showed that the proposed approach could effectively enhance the learning achievement of the students in comparison with the conventional context-aware learning system with single-stage prompts. Their study also indicated that more challenging tasks encouraged the students to put more effort into examining the contextual information and interpreting the learning content.

Therefore, in the teaching sequence of the progressive prompting strategy, each prompt will offer more information to learners than the previous one. That is, at the beginning, learners are given more abstract or simpler prompts to solve the questions. If they cannot solve the question, increasingly concrete prompts will be offered. In the meantime, the number of prompts depends on the ability of the student. The final prompt will be inclined to be a more detailed explanation of the question. Thus, it is important to offer proper prompts at the prime time to learners based on the learning activity.

The game-based learning approach with the progressive prompting strategy

In this study, we developed a computer game-based learning approach with a progressive prompting strategy in which students were given corresponding prompts progressively to help them figure out mathematics questions. At the beginning, the students were given more abstract or simpler prompts to solve the questions. If they could not solve the question, increasingly concrete prompts were offered. The final prompt was a detailed explanation of the question.

The mathematical game-based learning system adopted by the experimental group and control group was designed with the same learning content, using the digital game software, RPG Maker. The control group used the conventional math game, while the experimental group learned with the conventional math game and the extra two-tier progressive prompts. The learning content of the mathematical game-based learning system is divided into four maps, in the order of a forest, village, maze, and room. A total of 30 questions were answered by the learners involving two-step addition, two-step subtraction, and two-step addition and subtraction.

The mathematical game is named “Kingdom of Addition and Subtraction.” The protagonist is a warrior. In this kingdom, the king and queen gave birth to a beautiful and clever princess. Then one day, a terrible witch appeared and put the kingdom under a horrible curse. The curse put the people of the whole kingdom to sleep, with only the princess left awake. A fairy told the warrior of a rescue plan to save the kingdom. With the plot, the map, and the NPC (non-player character), the warrior must learn the concept of two-step addition and subtraction and answer the questions step by step. To complete the assigned task based on different levels, the
warrior must answer and get the antidote as well as bring enough money before they can be handed over to the princess. If the task is completed successfully, then the kingdom will be totally saved.

![Screenshot of two-step addition in the experimental group with progressive prompting](image1)

Figure 1. Screenshot of two-step addition in the experimental group with progressive prompting.

![Prompt with text description shown to players.](image2)

Figure 2. Screenshot of first-tier text description prompt.

In addition, an immediate feedback mechanism was set up in our system, so that when the learner chooses the wrong answer, the system will provide a prompt with the correct concept for relearning. In the experimental group, the progressive prompting mechanism was implemented. Figure 1 shows a math question delivered to the students on the screen: “A pencil costs 25 NT dollars, and correction fluid is more expensive than the pencil. A stapler is more expensive than the correction fluid by 13 NT dollars; please tell us how much the stapler cost.” If the students responded with the wrong answer, the learning system would give prompts with text description in the first-tier progressive prompting, as shown in Figure 2, then would give the learner the chance to answer the same question again. If the learner responded with the wrong answer again, the system would prompt with an image description in the second-tier progressive prompting, as shown in Figure 3. The learner is then given more chances to answer the same question until he/she has responded with the correct answer. Only in that case can the student learn the next concept. The framework of the progressive prompting educational game is shown in Figure 4. On the other hand, the feedback mechanism in the control group is the same as the first-tier prompting.
mechanism in the experimental group. Moreover, both groups were given exactly the same learning material, content, and assessments.

**Figure 3.** Screenshot of second-tier image description prompt

**Figure 4.** Framework of the progressive prompting educational game

**Experiment design**

To investigate the effect of game-based learning with the progressive prompting strategy, the second graders were invited to participate in the experiment. Their learning achievement, self-efficacy, and flow experience were examined.

**Participants**

There was a total of 58 second graders from two classes of an elementary school in northern Taiwan. Thirty students in the experimental group learned with the conventional math game with the extra two-tier progressive prompt, while 28 students in the control group learned with the conventional math game only. The two groups were taught by the same instructor for the experiment.

**Experimental process**

The purpose of this study was to guide the students to learn and enhance their learning outcomes by combining the characteristics of game-based learning and the progressive prompting strategy. The second graders
participated in the experiment of the “Two-step Addition and Subtraction unit” of the mathematics course in elementary school, as shown in Figure 5. Before conducting the experiment, students from both groups first spent 30 minutes filling in the questionnaires, including mathematics self-efficacy and math learning achievement as the pre-test scores. Then, the students were taught the two-stage concept in addition and subtraction for three classes (120 minutes). After the instruction, the experimental group engaged in game-based math learning with the progressive prompting strategy, while the control group learned with the conventional game-based math learning for 90 minutes. After the end of the learning activities, the two groups spent 30 minutes filling in questionnaires, including mathematics self-efficacy, flow experience, and mathematics learning achievement as the post-test scores to investigate whether the proposed approach affected the students’ learning achievement, self-efficacy, and flow experience.

**Figure 5. Experimental design of the study**

**Measuring tools**

The measuring tools employed in the study include questionnaires of flow experience, mathematics self-efficacy, and the pre-test and post-test of two-step addition and subtraction. Finally, the learners’ learning behavior patterns were recorded in the system.

The pre-test aimed to evaluate whether the two groups of students had an equivalent prior knowledge for learning the mathematics course. It consisted of three parts: four “simple addition” items (20 points), four “addition first and then subtraction” items (40 points), and five “subtraction” items (40 points) for evaluating the students’ prior knowledge. The post-test included four types of operation methods, two additions (20 points), addition first and then subtraction (30 points), subtraction first and then addition (30 points), and two subtractions (20 points). Moreover, in order to avoid resistance to the text reading comprehension by the second graders, the researchers designed multi-topic types including multiple choice questions, fill-in-the-blank questions, right and wrong, as well as column formulas. After the test questions were done, the mentor and two senior teachers with over 15 years of teaching experience were invited to review the learning achievement test. Therefore, the two-step addition and subtraction test had expert validity.

The self-efficacy scale is used to evaluate the level of self-efficacy when an individual performs a task. The scale was adapted on the basis of the Motivated Strategies Learning Questionnaire (MSLQ) developed by Pintrich et al. (1991). In line with the needs of the study, only the self-efficacy with seven questions from the Motivation questionnaire was involved and measured with a 5-point Likert scale. The reliability of the scale was tested by internal consistency with a Cronbach’s alpha of .87.

The flow experience scale mainly evaluates the state of learners who are fully engaged in the game-based learning, including both dimensions of pre-flow experience and flow experience, where the pre-flow experience dimension means the learners’ opinion based on the game-play experience, and the flow experience dimension
means the learners’ feeling during game-play. The study adopted the flow experience questionnaire from Wang and Chen (2010), including 22 items for pre-flow experience, and 12 for flow experience. A 5-point Likert scale was adopted in the study, ranging from 1 (strongly disagree) to 5 (strongly agree). The reliability of the scale was tested by internal consistency with pre-flow experience (Cronbach’s α = .92), flow experience (Cronbach’s α = .87), and the whole scale (Cronbach’s α = .94). The internal consistency coefficient is acceptable.

Coding scheme for lag-sequential analysis

One aim of the study was to investigate and trace the students’ learning behaviors in the game-based learning process through a code book related to learning behaviors, as seen in Table 1. This code book collected some codes in relation to two main classifications, namely irrelevant to the learning tasks and relevant to the learning tasks. This will help researchers understand the continuous relationship among the different behaviors and to look into the correlation among specific behaviors. All of the learning behaviors were automatically recorded and encoded in the database of the game-based learning system, and finally analyzed by lag-sequential analysis. In this study, the coding scheme for lag-sequential analysis was designed by three elementary school teachers who have rich teaching experience, and was examined by two professors of information education. Therefore, our coding scheme had expert validity.

<table>
<thead>
<tr>
<th>Code</th>
<th>Behavior</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>E</td>
<td>Exploration Dialogue with NPC or search for objects</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>Fighting Just fighting</td>
</tr>
<tr>
<td>V</td>
<td>V</td>
<td>Victory Win the fight</td>
</tr>
<tr>
<td>W</td>
<td>W</td>
<td>Withdraw Retreat from the fight</td>
</tr>
<tr>
<td>L</td>
<td>L</td>
<td>Losing Lost the fight</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>Transition Move to another scene</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Relevant to learning tasks</td>
</tr>
<tr>
<td>A</td>
<td>A</td>
<td>Answering Answer questions from NPC</td>
</tr>
<tr>
<td>H1</td>
<td>H1</td>
<td>1st step Get point from text description</td>
</tr>
<tr>
<td>H2</td>
<td>H2</td>
<td>2nd step Get point from image description</td>
</tr>
<tr>
<td>O</td>
<td>O</td>
<td>Task completed Correct answer to question asked</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>Task failed Wrong answer to question asked</td>
</tr>
</tbody>
</table>

Experimental results

This study investigated the effect of the proposed approach, game-based learning with a progressive prompting strategy, on the second graders’ learning achievement and self-efficacy, using one-way ANCOVA. If the test result reached a significant level ($p < .05$), then it represents a significant change. Finally, the Independent sample t test was used to explore whether the two groups of students had a change in the status of their flow experience. As for the students’ learning behaviors of the two groups, GSEQ can be used to present learners’ behavior patterns in the learning process.

Learning achievement

The Independent sample t test was employed to analyze the students’ learning achievement for the two groups before the experiment. The statistical result showed no significant difference ($t = 0.143$, $p = .887 > .05$) between the two groups for mathematics in the pre-test. Furthermore, one-way ANCOVA was used to analyze the effect on the students’ learning achievement for the experimental group and the control group after the experiment. The result showed that there was a significant difference ($F = 10.392$, $p = .002 < .01$) between the two groups, implying that students who learned with the game-based learning with progressive prompting strategy had superior learning performance compared with those who learned with the conventional game-based learning model without the progressive prompting strategy.

| Table 2. One-way ANCOVA results of the post-test scores for the experimental and control groups |
|-----------------------------------------------|----|-----|-------|-----|----|
| N                | Mean | SD  | Adjusted mean | F  |
| Progressive prompting game-based learning    | 30  | 87.70 | 12.74 | 87.44 | 10.39* |
| Conventional game-based learning              | 28  | 79.10 | 15.18 | 79.40 | 28  |

Note. *$p < .01$. 

328
Effect on the learning achievement for high and low self-efficacy students in the experimental group

To investigate whether there was any difference for students with high and low self-efficacy in the experimental group in terms of learning achievement, according to the score of the self-efficacy pre-test, the first 50% of students in the experimental group based on the math self-efficacy survey was classified as the high mathematics self-efficacy group (HMS), while the rest were classified as the low mathematics self-efficacy group (LMS). The Independent sample t test was employed to analyze the students’ learning achievement for the HMS and LMS of the experimental group before the experiment. The statistical result showed that there was significant difference \((t = 2.09, p = .04 > .05)\) between the HMS and LMS before the learning activity. However, there was no significant difference \((t = 0.96, p = .45 > .05)\) between the HMS and LMS after the two-step addition and subtraction unit, implying that the proposed approach could gradually scaffold LMS students in math learning process so that shorten the difference between HMS and LMS students in mathematical learning achievement, as shown in Table 3.

Table 3. Independent sample t test of learning performance for students with high and low mathematics self-efficacy in the experimental group.

<table>
<thead>
<tr>
<th>Items</th>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test</td>
<td>HMS</td>
<td>16</td>
<td>88.56</td>
<td>9.18</td>
<td>2.09</td>
<td>.04</td>
</tr>
<tr>
<td></td>
<td>LMS</td>
<td>14</td>
<td>75.63</td>
<td>11.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post-test</td>
<td>HMS</td>
<td>16</td>
<td>88.98</td>
<td>17.02</td>
<td>0.96</td>
<td>.45</td>
</tr>
<tr>
<td></td>
<td>LMS</td>
<td>14</td>
<td>86.24</td>
<td>11.74</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .05. High math self-efficacy group (HMS), Low math self-efficacy group (LMS).

Results of flow experience

To investigate the effect of the proposed approach on the students’ flow experience for the two groups, the students in the experimental group and control group filled in the questionnaire of Flow experience after the learning activity. As shown in Table 4, there were no significant differences in either dimension of pre-flow experience or flow experience for the two groups through analysis of the Independent t test \((p > .05)\). This result implies that the students of the experimental group and the control group had the same flow experience whether or not they were using the progressive prompting strategy. That is, the students in the experimental group did not change their flow experience when the progressive prompting strategy was added to the math game-based learning activity.

Table 4. Statistical results of the Independent t test of flow experience for the two groups

<table>
<thead>
<tr>
<th>Items</th>
<th>Experimental group ((n = 30))</th>
<th>Control group ((n = 28))</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Pre-flow experience</td>
<td>4.26</td>
<td>0.60</td>
<td>4.16</td>
<td>0.52</td>
</tr>
<tr>
<td>Flow experience</td>
<td>4.13</td>
<td>0.64</td>
<td>4.06</td>
<td>0.55</td>
</tr>
</tbody>
</table>

Analysis of students’ learning behaviors

To further investigate the effect of the proposed approach on the students’ learning achievement, lag-sequential analysis was employed to explore the students’ learning behaviors for the experimental group and the control group. The lag-sequential analysis method aims to find the significant learning patterns of the coded behaviors. To ensure the quality of the coding process, the consistency of the two researchers’ coding was examined. It was found that the inter-rater Kappa reliability of the two researchers’ coding was .82 for the experimental group and .80 for the conventional game-based learning group, demonstrating high reliability. After performing the lag-sequential analysis, a Z-score was obtained to represent significance between each pair of the coded behaviors (Chiang, Yang, & Hwang, 2014; Hou et al., 2009). If the Z-score was greater than 1.96, it was concluded that the two behaviors had a significant sequential relationship in the learning context (Bakeman & Gottman, 1997).

Tables 5 and 6 show the Z-score values of the learning behaviors of the experimental group and control group. The behavioral codes listed in the left-most column were the starting behaviors, while those listed in the top-most row were the resulting behaviors. The value in each entry of the tables was the Z-score. The significant relationship was marked with a “*” if the Z-score was greater than 1.96. For example, the sequential relationship
“H2(stage 2 progressive prompting)”→“O(Answer correct)” was significant since the corresponding Z-score was 2.87.

Table 5. Z-scores for the behaviors of the students who learned with the progressive prompting game-based learning

<table>
<thead>
<tr>
<th>Given code</th>
<th>A</th>
<th>H1</th>
<th>H2</th>
<th>O</th>
<th>X</th>
<th>E</th>
<th>F</th>
<th>V</th>
<th>W</th>
<th>L</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-22.73</td>
<td>32.03</td>
<td>-10.79</td>
<td>32.45</td>
<td>-4.47</td>
<td>-11.73</td>
<td>-9.28</td>
<td>-2.11</td>
<td>-1.49</td>
<td>-8.93</td>
<td>-9.74</td>
</tr>
<tr>
<td>H1</td>
<td>-11.77</td>
<td>-6.23</td>
<td>44.03</td>
<td>2.09</td>
<td>-2.34</td>
<td>-6.14</td>
<td>-4.94</td>
<td>-1.11</td>
<td>-0.78</td>
<td>-4.68</td>
<td>-5.18</td>
</tr>
<tr>
<td>H2</td>
<td>-10.79</td>
<td>-5.65</td>
<td>-5.12</td>
<td>21.76</td>
<td>25.35</td>
<td>-5.57</td>
<td>-4.48</td>
<td>-1.00</td>
<td>-0.71</td>
<td>-4.24</td>
<td>-4.69</td>
</tr>
<tr>
<td>O</td>
<td>35.84</td>
<td>-11.89</td>
<td>-10.63</td>
<td>-22.67</td>
<td>-4.46</td>
<td>-3.36</td>
<td>4.83</td>
<td>-2.11</td>
<td>-1.49</td>
<td>-8.92</td>
<td>11.95</td>
</tr>
<tr>
<td>X</td>
<td>-4.47</td>
<td>-2.34</td>
<td>25.35</td>
<td>-4.46</td>
<td>-0.88</td>
<td>-2.31</td>
<td>-1.85</td>
<td>-0.42</td>
<td>-0.29</td>
<td>-1.76</td>
<td>-1.94</td>
</tr>
<tr>
<td>E</td>
<td>18.72</td>
<td>-6.14</td>
<td>-5.57</td>
<td>-11.71</td>
<td>-2.31</td>
<td>-0.87</td>
<td>-4.15</td>
<td>-1.09</td>
<td>-0.77</td>
<td>-4.61</td>
<td>11.26</td>
</tr>
<tr>
<td>V</td>
<td>-0.16</td>
<td>-1.11</td>
<td>-1.00</td>
<td>-2.11</td>
<td>-0.42</td>
<td>-1.09</td>
<td>10.07</td>
<td>-0.20</td>
<td>-0.14</td>
<td>-0.83</td>
<td>-0.92</td>
</tr>
<tr>
<td>W</td>
<td>1.26</td>
<td>-0.78</td>
<td>-0.71</td>
<td>-1.49</td>
<td>-0.29</td>
<td>-0.77</td>
<td>2.82</td>
<td>-0.14</td>
<td>-0.10</td>
<td>-0.59</td>
<td>1.00</td>
</tr>
<tr>
<td>L</td>
<td>0.98</td>
<td>-4.68</td>
<td>-4.24</td>
<td>-8.92</td>
<td>-1.76</td>
<td>-4.61</td>
<td>32.12</td>
<td>-0.83</td>
<td>-0.59</td>
<td>-3.51</td>
<td>3.10</td>
</tr>
<tr>
<td>T</td>
<td>-9.59</td>
<td>-5.18</td>
<td>-4.69</td>
<td>-9.88</td>
<td>-1.94</td>
<td>51.13</td>
<td>-4.10</td>
<td>-0.92</td>
<td>-0.65</td>
<td>-3.89</td>
<td>-4.30</td>
</tr>
</tbody>
</table>

Note. *Z > 1.96.

Table 6. Z-scores for the behaviors of the students who learned with the conventional game-based learning

<table>
<thead>
<tr>
<th>Given code</th>
<th>A</th>
<th>H1</th>
<th>H2</th>
<th>O</th>
<th>X</th>
<th>E</th>
<th>F</th>
<th>V</th>
<th>W</th>
<th>L</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>O</td>
<td>15.89</td>
<td>-11.55</td>
<td>-10.65</td>
<td>0.56</td>
<td>2.82</td>
<td>-2.00</td>
<td>-1.22</td>
<td>-5.02</td>
<td>7.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>9.08</td>
<td>-5.95</td>
<td>-5.48</td>
<td>-2.69</td>
<td>-1.06</td>
<td>-1.03</td>
<td>-0.63</td>
<td>-3.05</td>
<td>8.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>-1.04</td>
<td>-2.01</td>
<td>-1.85</td>
<td>-1.03</td>
<td>5.92</td>
<td>-0.35</td>
<td>-0.21</td>
<td>1.11</td>
<td>1.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W</td>
<td>-1.06</td>
<td>-1.23</td>
<td>-1.13</td>
<td>-0.63</td>
<td>4.23</td>
<td>-0.21</td>
<td>-0.13</td>
<td>1.07</td>
<td>1.34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>-4.55</td>
<td>-5.92</td>
<td>-5.46</td>
<td>-3.05</td>
<td>27.25</td>
<td>-1.03</td>
<td>-0.63</td>
<td>-2.66</td>
<td>6.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>-7.15</td>
<td>-5.28</td>
<td>-4.86</td>
<td>39.03</td>
<td>-2.84</td>
<td>-0.91</td>
<td>-0.56</td>
<td>-2.70</td>
<td>-2.39</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *Z > 1.96.

Figure 6. Students’ behavioral sequences for the experimental group

A total of 3,743 learning patterns were recorded for the experimental group (N = 30), while 2,148 for the control group (N = 28) were collected. Figure 6 shows the behavioral sequence for the students in the experimental group, A→H1(Answering→1st step), H1→H2 (1st step→2nd step), H2→O (2nd step→ Task completed), H2→X (2nd step→ Task failed), sequentially, indicating that it is significant to provide prompts to the students in the experimental group, but the students could give right answers or wrong answers after the second prompt was provided. Figure 7 shows the behavioral sequence for the students in the control group, A→X (Answering→Task
failed), X→A (Task failed→Answering), sequentially, indicating that even if the answer was wrong, the students would try to continually answer. As we can see from Figure 6 and Figure 7, the students from the two groups could go all out to complete the learning tasks. It is particularly noteworthy that the students in the experimental group who presented a significant behavioral pattern from prompting to the answering process through the progressive prompts were able to complete the questions and gain better learning achievement.

**Figure 7. Students’ behavioral sequences for the control group**

**Discussion and conclusions**

This study investigated the effects of a game-based learning approach with a progressive prompting strategy on second graders’ mathematics learning achievements related to “two-step addition and subtraction,” as well as mathematics self-efficacy and flow experience. A computer educational game has been developed accordingly. An experiment has been conducted on an elementary school mathematics course to evaluate the performance of the proposed approach. In addition, students’ learning patterns were analyzed via using the GSEQ method to further understand their learning process. The research findings show that the proposed approach could enhance students’ learning achievement more than the conventional game-based learning approach, showing that students can acquire the concept from the first prompt with text description; if they cannot understand the abstract text, they will receive the feedback with more concrete content from the second prompt with image description. After this stage, the learners are able to construct the integrated mathematical image. In other words, the progressive prompting mechanism brings a massive benefit for concept learning. Moreover, the proposed approach would not affect the students’ flow experience in game playing. That is, a game-based learning approach with a progressive prompting strategy with learning activities has been shown to be beneficial to students’ mathematics improvement. From these findings, the research questions are answered as follows:

- The students who used the game-based learning approach with the progressive prompting strategy showed significantly better learning achievements than those who used the conventional game-based learning approach.
- In the beginning, the learning achievements of the high mathematics self-efficacy group (HMS) were significantly better than those of the low mathematics self-efficacy group (LMS) in the experimental group. After the learning activity, there was no significant difference between the HMS and LMS groups. It can be inferred that the proposed approach can effectively reduce the difference in the mathematical learning achievement of HMS and LMS students.
- The students who learned the game-based learning approach with the progressive prompting strategy and the conventional game-based learning approach showed equivalent flow experiences.
- According to the analytic results of the students’ learning behaviors, those who learned with the proposed approach can figure out the questions successfully after they receiving the progressive prompts. On the other hand, those who learned with the conventional game-based learning approach revealed some difficulties in the learning process.
In recent years, analysis of learning behaviors has attracted the attention of researchers in the study of game-based learning (Hou, 2015; Hwang, Hsu, Lai & Hsueh, 2017). Behavioral analysis helps learners analyze their learning behaviors in the learning process and investigate the attribution of research results. The research findings show that the students in the experimental group presented a significant behavioral pattern from prompting to the answering process through the progressive prompts which could help guide them to complete the questions, improve their critical thinking, and finally gain better learning achievement. On the other hand, those in the control group needed to go through a trial and error process, whereby the students kept trying to answer the questions and failing. Once they correctly answered the question, they might not really understand the correct steps to solve questions.

As for their mathematics self-efficacy, the proposed approach could effectively reduce the difference in the mathematical learning achievement of the high mathematics self-efficacy group (HMS) and low mathematics self-efficacy group (LMS) students of the experimental group; that is, the proposed approach could help LMS students understand the learning content, and improve their learning achievement. This might be the reason why the LMS students could solve the questions based on correct prompts. As suggested by Vygotsky (1978), it is important to provide prompt supports to students when they face difficulties in the learning process. We inferred that it is more difficult for the LMS learners to imagine the situation during mathematics problem solving, so it causes lower self-efficacy; however, with the progressive prompting mechanism, it could assist students to imagine the concept, thus reducing the difference with the HMS learners. Although the proposed approach has been shown to enhance students’ learning achievement, there are still some limitations to this study. First of all, only a 90-minute learning activity was designed in the experiment. If the learning design included more experimental learning activities, the researchers could further explore the relationship between meta-cognition and learning achievement in GBL.

The research findings also show that there was no significant difference between the experimental group and control group in terms of flow experience. Previous studies have indicated that the GBL approach could improve learners’ flow experience (Chang et al., 2012; Hwang et al., 2015). In this study, two groups were given GBL to learn mathematics, so it is reasonable that there was no difference between the two groups, implying that the students were happy to get involved in the activity. It also indicates that the progressive prompting strategy designed for the experimental group is appropriate and will not affect students’ learning. Hwang, Chu, Yin, and Ogata (2015) emphasized that an enjoyable and challenging game accompanied with appropriate learning strategies could help learners achieve effective learning achievement and learning goals. Therefore, the study incorporated the progressive prompting strategy into the second graders’ mathematics game-based learning context.

In summary, this study validates four points. First, the educational game with the progressive prompting mechanism which provides learner multi-tier feedback from abstract to concrete hints could assist the learning of concepts, and increase learning achievement; second, with the concrete image description prompt, the proposed method could effectively assist low self-efficacy learners to construct the concrete image of mathematics, then decrease the difference with high self-efficacy students; third, based on the result of flow experience, we could say that combining the progressive prompting mechanism into game-based learning may not influence learners’ feeling during game-playing; finally, the results of learning behavior analysis shows that the proposed method could help learners not only respond questions, but also to learn the math concepts, and gradually consolidate a stable foundation of mathematics.

Acknowledgments

This study is supported in part by the Ministry of Science and Technology of the Republic of China under contract number MOST 105-2511-S-152-005-MY2, MOST 104-2511-S-031 -003 -MY4 and MOST 106-2511-S-031-001-MY3.

References


