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 The articles should be original, unpublished, and not in consideration for publication elsewhere at the time of submission to Educational Technology & Society and three months thereafter.

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


Founding Editor
Kinshuk, 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, Winson-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 Jendi, University of Tunis, Tunisia; Mike Joy, University of Warwick, United Kingdom; Athanasis Karoulis, Hellenic Open University, Greece; Paul Kirchner, 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 Leuliette, Universite Laval, Canada; Tzu-Chien Liu, National Central University, Taiwan; Rory McCreath, Athabasca University, Canada; David Merrill, Brigham Young University - Hawaii, USA; Marcelo Milrad, Vaxjo University, Sweden; Riichiro Mizoguchi, Osaka University, Japan; Permanand Mohan, The University of the West Indies, Trinidad and Tobago; Kiyoji 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. Shih, 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 Teo, Nanyang Technological University, Singapore; Chin-Chung Tsai, National Taiwan Normal University, Taiwan; Jei Chi Yang, National Central University, Taiwan; Stephen J. H. Yang, National Central University, Taiwan; Yu-Mei Wang, University of Alabata at Birmingham, USA; Ashok Patel, CAL Research & Software Engineering Centre, UK; Reinhard Oppermann, Fraunhofer Institut Angewandte Informationstechnik, Germany; Vladimir A Fominich, K. E. Tsiolkovsky Russian State Tech Univ, Russia; Olga S Fominnova, Studio “Culture, Ecology, and Foreign Languages,” Russia; Piet Koomers, University of Twente, The Netherlands; Chul-Hwan Lee, Inchon National University of Education, Korea; Brent Muirhead, University of Phoenix Online, USA; Erkki 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; Wei-Chich Fang, National Sun Yat-sen University, Taiwan.

Technical Manager
Jie Peng, National Taiwan Normal University, Taiwan

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, He ping E. Rd., Education Building, Rm. 611, Taipei 10664, Taiwan.

ISSN 1436-4532 (online) and 1176-3647 (print). This article of the Journal of Educational Technology & Society is available under Creative Commons CC-BY-ND-NC 3.0 license (https://creativecommons.org/licenses/by-nd-nc/3.0). For further queries, please contact Journal Editors at ets.editors@gmail.com
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.E.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.ets.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).
## Table of contents

### Full Length Articles

- **Comparison of Animation and Static-picture based Instruction: Effects on Performance and Cognitive Load for Learning Genetics**  
  Chi Yan, Chun-Hui Jen, Chun-Yen Chang and Ting-Kuang Yeh  
  1–11

- **Effects of Multimodal Learning Analytics with Concept Maps on College Students’ Vocabulary and Reading Performance**  
  Shih-Ping Wang and Yih-Lan Chen  
  12–25

- **Developing a Multidimensional Framework for Analyzing Student Comments in Wikis**  
  Xiao Hu, Christy Weng-Lam Cheong and Samuel Kai-Wah Chu  
  26–38

- **Strengthening Social Networks in Online Discussion Forums to Facilitate Help Seeking for Solving Problems**  
  Po-Yao Chao, K. Robert Lai, Chen-Chung Liu and Hung-Ming Lin  
  39–50

- **Analysing Group Dynamics of a Digital Game-based Adventure Education Course**  
  Chang-Hsin Lin and Ju-Ling Shih  
  51–63

- **Social Networking Sites as Formal Learning Environments in Business Education**  
  Abida Ellahi  
  64–75

- **Factors Related to ICT Competencies for Students with Learning Disabilities**  
  Ting-Fang Wu, Cheng-Ming Chen, Hui-Shan Lo, Yao-Ming Yeh and Ming-Chung Chen  
  76–88

- **Mediating Effects of Individuals’ Ability Levels on the Relationship of Reflective-Impulsive Cognitive Style and Item Response Time in CAT**  
  Chao Wang and Hong Lu  
  89–99

- **Increasing Information Reposting Behavior in Online Learning Community**  
  Omid R. B. Speily and Ahmad A. Kardan  
  100–110

### Guest Editorial

- **Authentic Edutainment with Advanced Technologies**  
  Rustam Shadiev, Wu-Yuin Hwang, George Ghinea and Maiga Chang  
  111–114

### Special Issue Articles

- **Development of a SoLoMo Game-Based Application for Supporting Local Cultural Learning in Taiwan**  
  Yen-Ting Liu, Yu-Ming Tseng, Yi-Sheng Lee, Tz-Chi Wang, Szu-I Tsai and Yan-Jhih Yi  
  115–128

- **User-Oriented EFL Speaking through Application and Exercise: Instant Speech Translation and Shadowing in Authentic Context**  
  Thi-Huyen Nguyen, Wu-Yuin Hwang, Xuan-Lam Pham and Zhao-Heng Ma  
  129–142

- **Application and Analysis of a Mobile E-Book System Based on Project-Based Learning in Community Health Nursing Practice Courses**  
  Ting-Ting Wu, Yueh-Min Huang, Chen-Ying Su, Lei Chang and Yi Chen Lu  
  143–156

- **Enhanced Agility of E-Learning Adoption in High Schools**  
  Gebremaniam Mesfin, Gheorghe Ghinea, Tor-Morten Grønli and Wu-Yuin Hwang  
  157–170

- **Educational Games to Enhance Museum Visits for Schools**  
  Benoît Bossavit, Alfredo Pina, Isabel Sanchez-Gil and Aitziber Urtasun  
  171–186

- **Using Exaggerated Feedback in a Virtual Reality Environment to Enhance Behavior Intention of Water-Conservation**  
  Wei-Che Hsu, Ching-Mei Tseng and Shih-Chung Kang  
  187–203
Gender-Related Differences in Collaborative Learning in a 3D Virtual Reality Environment by Elementary School Students
Yi-Lien Yeh, Yu-Ju Lan and Yen-Ting R. Lin

A Study of the Use of Wearable Devices for Healthy and Enjoyable English as a Foreign Language Learning in Authentic Contexts
Rustam Shadiev, Wu-Yuin Hwang and Tzu-Yu Liu

Investigating Flipped Classroom and Problem-based Learning in a Programming Module for Computing Conversion Course
Adriana Elena Chis, Arghir-Nicolae Moldovan, Lisa Murphy, Pramod Pathak and Cristina Hava Muntean

Abstract: Game-driven Keyword Auction and Summarization for Academic Reading
Hercy N. H. Cheng, Calvin C. Y. Liao and Wan-Chen Chang

Pedagogical Change in Mathematics Learning: Harnessing the Power of Digital Game-Based Learning
Siew Pei Hwa

Gamifying and Mobilising Social Enquiry-based Learning in Authentic Outdoor Environments
Morris Siu-Yung Jong, To Chan, Ming-Tak Hue and Vincent W. L. Tam
Comparison of Animation and Static-picture based Instruction: Effects on Performance and Cognitive Load for Learning Genetics

Chi Yang1, Chun-Hui Jen1, Chun-Yen Chang1,2* and Ting-Kuang Yeh1,2,3*

1Science Education Center and Graduate Institute of Science Education, National Taiwan Normal University, Taiwan // 2Department of Earth Sciences, National Taiwan Normal University, Taiwan // 3Institute of Marine Environmental Science and Technology, National Taiwan Normal University, Taiwan // chyi51757@gmail.com // jen.chunhui@gmail.com // changey@ntnu.edu.tw // tkyeh@ntnu.edu.tw

(Submitted November 30, 2016; Revised March 4, 2017; Accepted April 28, 2017)

ABSTRACT

This study aimed to examine the relative effectiveness of using an animation versus static pictures in terms of supporting the learning of genetics. To provide a methodologically sound comparison, the two sessions were constructed to be equivalent and were designed based on principles provided by the cognitive theory of multimedia learning and the cognitive load theory. Genetics was chosen as the instructional content, which included learning about the processes of cell division, mitosis, and meiosis. The results indicate that students in the animation group perceived less extraneous cognitive load and achieved a better learning outcome than those in the static pictures group. Therefore, this study supports the superiority of the animation over static picture instruction when learning micro-scientific phenomena.

Keywords
Multimedia learning, Genetics learning, Computer animation, Cognitive load, Static pictures

Introduction

The integration of multimedia technologies to promote learners’ cognitive development is considered one of the most important objectives for science education (Mayer & Moreno, 2003; Yeh et al., 2012). Several studies have provided evidence that multimedia instructional environments can help students learn more effectively than traditional strategies (Jereb & Smitek, 2006; Najjar, 1996). However, the application of appropriate instruments to the design of effective multimedia-based educational tools has not been studied sufficiently (Mayer, 2008; Tabbers, Martens, & van Merrienboer, 2004). One important long-debated issue is the effectiveness of using computer animation and static pictures to support students’ learning (Höffler & Leutner, 2007).

It is often assumed that animation can help learners mentally visualize a dynamic process compared to static pictures; however, this assumption is still controversial. Researchers have believed that animation visualization offers a realistic representation of the to-be-explained processes, facilitates a deeper comprehension of dynamic systems and inspires greater motivation to learn (Höffler & Leutner, 2007; Mayer, Hegarty, Mayer, & Campbell, 2005). In contrast, it is also believed that some possible cognitive advantages are associated with learning from static picture instruction. For example, (1) static pictures reduce cognitive processes because learners see only frames that represent key steps in the process, helping them focus on the most relevant information, and (2) learners are encouraged to explain the changes from one frame to the next, thereby enhancing active information processing (Mayer et al., 2005).

There are no consistent findings regarding the superiority of animation or static picture instruction in learning processes. A number of studies have reported that animated graphics can be used more effectively than static graphics to illustrate difficult abstract concepts and to visualize dynamic processes (Bodemer, Ploetzner, Feuerlein, & Spada, 2004; Williamson & Abraham, 1995). However, animation-based approaches are not always advantageous over static pictures. In a study conducted by Starbek, Starčič Erjavec, and Peklaj (2010), four groups of high-school students were taught the process of protein synthesis through a traditional lecture, text only, animation, and static pictures. In this study, Starbek and colleagues (2010) found that the animation group and the static picture group acquired better knowledge and improved their comprehension skills compared to the other two groups. However, the animation group, and the static picture group did not differ from each other, which failed to confirm the superior impact of animation. A similar finding was also reported by Rieber (1989) in a sample of 192 high school students; no significant differences in learning achievement were found between the graphic and animation groups.

To clarify the superiority of animation over the static pictures approach, in the following paragraphs, we discuss two crucial points that should be taken into account but were usually overlooked in previous studies. The first
point concerns the methodology issue, which focuses on the equivalent design between animations and static picture instruction; the second point concerns the instructional design, emphasizing the application of learning theories to reduce learners’ cognitive load.

It is vital to reduce potential non-equivalence problems by controlling variables other than the presentation format when comparing the relative effectiveness of animation/static pictures. By doing so, the argument can be made that the differences in the outcomes result from inherent differences between the formats (animation versus static pictures) rather than the poor design of either of these. However, after reviewing the literature, we found that there are at least three sources of non-equivalence between instructions that will lead to incomparability and render conclusions difficult. The first one is the amount of information provided by instructions. For example, to compare the relative effectiveness of different instructions, Starbek et al. (2010) provided participants with static illustrations and text to summarize protein synthesis in one experimental condition, and provided two animations showing the transcription and translation processes of protein synthesis, which were supplemented with English text in another condition. In this case, the two conditions not only differed in the dynamics of the materials but also in the graphic and the text information the materials provided. The second possible source of non-equivalence is the interface of the instructions. For example, in Mayer et al. (2005), the authors compared the relative effectiveness between computer-based animation and paper-based static diagrams, and found that static picture instruction led to better learning compared with animation. Although Mayer et al. attempted to maintain the informational equivalence between the two instructions, the difference in the interface between the two conditions (i.e., computer-based vs. paper-based) still confounded the results. The third possible source of non-equivalences is interactivity. Computer-based instructions are generally designed with functions that allow learners to interact with the instructions. However, in some studies, interactivity was not treated as a controlling variable when comparing the relative effectiveness between animations and static pictures. For example, in Watson, Butterfield, Curran, and Craig (2010), animations included self-pacing functions that were compared with static instructions, which presented no interaction. On the contrary, in Ayres, Marcus, Chan, and Qian (2009), learners had to use the scroll function to observe the information they needed in static picture instruction, but this function was not included in the animation presentations. Although Watson et al. and Ayres et al. both found that the animation condition was superior to the static picture condition since the interactivity of the instructions also varied with the presenting format in their studies, the influence of presentation format on learning outcomes is ambiguous.

It has been indicated that the design of multimedia instructions should follow principles that help to lower learners' cognitive load. For example, Ayres and Paas (2007) argued that many animated instructional environments are not as effective as expected because they create an unnecessarily large cognitive load. Similarly, Tversky, Morrison, and Betrancourt (2002) have indicated that dynamic visualization is no more effective than static visualization if the animation is too complex or the material is presented too quickly. Therefore, Ayres and Pass (2007) as well as Tversky et al. (2002) suggested that the design of multimedia instructions must be based on well-established principles that help to reduce learners' cognitive load through careful control of the manner of presentation and by attracting learners' attention in a more appropriate way. Although researchers acknowledge the importance of applying well-established principles in the design of effective multimedia instruction, not all studies have clearly mentioned on what principles or theories the design was based.

To look more carefully into the superiority of animation in comparison with the static picture approach, in this study, we took the equivalence issue into account and applied well-established principles in the design of instructions. To keep the design equivalent, both animation and the static picture instruction (1) provided identical information for learners to learn, (2) were designed as computer-based instructions, and (3) were designed with interactive functions (i.e., self-pacing). To ensure that the instructions were well-designed to reduce learner’s cognitive load, both instructions were designed based on principles suggested by the cognitive load theory (CLT; Sweller, Ayres, & Kalyuga, 2011; Sweller, Chandler, Tierney, & Cooper, 1990; Sweller, Van Merrienboer, & Paas, 1998) and the cognitive theory of multimedia learning (CLMT; Mayer, 2001; Mayer, 2003; Mayer & Moreno, 2002). Both CLT and CTML propose minimizing unnecessary cognitive load, which is the central consideration in the design of multimedia materials, and provide research-based principles that could be applied in the design of multimedia materials to reduce learners’ cognitive load. By applying principles suggested by CLT and CLTM, the instructions used in this study were designed to reduce the processing of extraneous materials and to allow learners to deeply study the materials.

This study specifically focused on comparing an animation of invisible infinitesimal phenomena with an equivalent static picture instruction. Hegarty (2004) has argued that researchers must go beyond making a simple distinction between static and dynamic displays because the latter could be further categorized as (1) animations
of visible phenomena (e.g., a machine in motion), (2) animations of invisible phenomena (e.g., changes in temperature on a weather map), and (3) animations of abstract information (e.g., statistical concepts). Therefore, the results from studies with one type of dynamic display may not inevitably generalize to others. To respond to what Hegarty (2004) asserted, we have paid attention to the superiority of animations of invisible infinitesimal phenomena since it has been suggested that animations may be more effective when they are utilized to visualize invisible phenomena in the real world (Narayanan & Hegarty, 2002).

Genetics was chosen as the learning topic in this study. With recent advances in gene technology and the rapid progress on the Human Genome Project, genetics is becoming an important learning prerequisite in high school curricula. However, since the principles of genetics are abstract and the phenomena of genetics are infinitesimal and cannot be observed by the naked eye, students often develop misconceptions and become disoriented while studying this topic. Studies have found that students are weak in learning genetic terminology (Albaladejo & Lucas, 1988; Banet & Ayuso, 2003) and the invisible infinitesimal processes of cell division or meiosis (Brown, 1990; Stewart, Hafner, & Dale, 1990). The difficulty in learning genetics not only puzzles students but also poses a large challenge to most genetics teachers (Johnstone & Mahmoud, 1980). In Taiwan, the topic of genetics is covered extensively in the subjects of Science and Life Technology Curriculum Standards (Grades 1–9) and Life Science Curriculum Guidelines (Grades 10–12). Given that the understanding of genetics is an integral component of science education, it is crucial to design appropriate instructional tools to facilitate learning.

This study thus aimed to examine the relative effectiveness of using animation versus static pictures in terms of supporting the learning of invisible infinitesimal scientific phenomena. In this article, “relative effectiveness” refers to an instruction (1) that has less cognitive load and/or (2) that helps learners to reach higher achievement compared to the other instruction. Therefore, the investigation was guided by answering the following questions:
- Whether animations of invisible infinitesimal phenomena require less cognitive load than equivalent static picture instruction.
- Whether animations of invisible infinitesimal phenomena help learners to reach higher achievement than equivalent static picture instruction.

**Methods**

**Participants**

A total of 181, 7th grade students from six classes at a public junior high school were recruited and were assigned to one of two groups based on their class. Three of the six classes were assigned to the static pictures group ($n = 89$), and the remaining three classes were assigned to the animation group ($n = 92$). According to the data collected from the Intelligence Test for Junior High School--Revised Edition (Chen, 2004), these two groups did not differ from each other in their verbal reasoning, mathematics reasoning, and figural reasoning abilities.

**Procedure**

The study procedure consisted of a prior knowledge estimation phase, a learning phase, and a testing phase in sequence. In the prior knowledge estimation phase, students’ prior knowledge about genetics concepts was assessed. In the learning phase, students attended two 50-minute class-periods with multimedia instruction followed by an immediate evaluation of the extent of cognitive load they perceived during the curriculum. The content of the instruction was identical in both groups, but different media (i.e., static pictures versus animation) were used. In the testing phase, students were asked to complete an achievement test to evaluate their learning outcomes.

**Designs of the multimedia instructions for learning genetics**

The instructions were designed to promote an understanding of abstract concepts and invisible infinitesimal phenomena in genetics by applying principles of multimedia design suggested by CLT and CTML. Both animation and static picture instruction presented important concepts and phenomena in genetics, including (1) genetic material, (2) the process of cell division, (3) the process of mitosis, and (4) the process of meiosis. These four topics were sequentially presented in four segments. Animation instruction and static picture instruction were both computer-based, designed with self-pacing functions, and provided identical information for learners.
to learn, except that the animation was transitional in nature whereas the static pictures were not. To generate the content of the two sessions, we listed the main stages of the dynamic process to be learned (e.g., the process of mitosis) first, and then created corresponding diagrams and text to illustrate each stage. Static picture instruction thus was composed of a series of frames that presented the illustrations we created on each. Animation instruction was generated based on the frames used in the static picture instruction, but was designed to present continuously the minute changes occurring in the process.

The instructions were designed to reduce the processing of extraneous materials. The materials were presented as animation/static pictures and text together. Text was placed next to target images to reduce the split-attention effect (Sweller et al., 2011; see Figure 1). We avoided presenting materials irrelevant to basic concepts or essential elements (i.e., the coherence principle; Mayer & Moreno, 2003). Moreover, we highlighted key words and important concepts to help learners extract the main concepts and organize the information efficiently (i.e., the signaling principle; see Figure 2) (Mayer & Moreno, 2003).

The instructions were designed to allow learners to consider the materials deeply. To help learners to process the multimedia materials more efficiently, we produced a pre-training section for each topic and presented it before the illustrations, in which we briefly introduced glossaries and important concepts (i.e., pre-training principle; see Figure 2) (Mayer & Moreno, 2003). Since learners usually do not have sufficient time and capacity to handle animation material, especially when the content is rich and complex or the speed of the presentation is too fast to realize the context, the instructions were divided into meaningful segments, and learners could decide to watch the next or previous segment by clicking arrowhead indicators on the screen (i.e., segmentation principle; Mayer & Moreno, 2003; see Figure 1). Two levels of segments were constructed in the instructions. The first-level segments were the four topics to be learnt; each first-level segment was composed of a pre-training segment and an illustration segment. Moreover, the instructions were designed to allow learners to control the pace of the presentation. For example, in animation instruction, learners could decide the pace of the presentation by dragging a scrollbar back and forth or clicking a pause or play button (Figure 1). Similarly, the presentation of static pictures allowed learners to determine whether to go to the next stage, the previous stage, or back to the first stage by choosing the corresponding button (Figure 3).

Figure 1. An example of the instructional section in animation instruction
Figure 2. An example of the instructional section

Note. Translated content: Mitosis Overview: Mitosis is a process of cell division that results in two identical daughter cells. Mitosis is divided into four stages: prophase, metaphase, anaphase and telophase. Each stage has its own process. Before mitosis, to prepare for the process of cell division, centrioles and chromosomes are duplicated. After mitosis, each daughter cell will have the same number of chromosomes as the parent cell.

Figure 3. An example of the instructional section in the static picture instruction

Instruments

Three instruments were used in this study to assess students’ prior knowledge, to evaluate a subjective index of the cognitive load of the genetics curriculum and to assess students’ achievements regarding genetic concepts. Sample instrument questions are shown in Table 1.
Table 1. Sample items of instruments used in this study

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Sample items</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFT</td>
<td>• Which one is a characteristic of the nucleus?</td>
<td>Multiple-choice</td>
</tr>
<tr>
<td></td>
<td>(a) There is chloroplast in the nucleus for photosynthesis;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(b) It contains hereditary (genetic) information and is the life center of</td>
<td></td>
</tr>
<tr>
<td></td>
<td>a cell;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(c) It supports the cell skeleton, preventing the cell from becoming deformed;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(d) It is a place for energy production and nutrient oxidation.</td>
<td></td>
</tr>
<tr>
<td>CLQ</td>
<td>• Substantial mental effort was required to understand the course content of</td>
<td>7-point scale</td>
</tr>
<tr>
<td></td>
<td>mitosis. (overall cognitive load)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• The text font, size and color presented took substantial mental effort to</td>
<td></td>
</tr>
<tr>
<td></td>
<td>read. (extraneous cognitive load)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Substantial mental effort was required to operate the computer. (extraneous</td>
<td></td>
</tr>
<tr>
<td></td>
<td>cognitive load)</td>
<td></td>
</tr>
<tr>
<td>GCT</td>
<td>• Please describe what you know about chromosomes, DNA, and genes and the</td>
<td>Open-ended</td>
</tr>
<tr>
<td></td>
<td>relationship among them.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• A creature has 10 pairs of chromosomes in its cells. After five successive</td>
<td>Multiple-choice</td>
</tr>
<tr>
<td></td>
<td>mitotic divisions, how many pairs of chromosomes does each daughter cell have?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(a) 50 pairs;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(b) 2 pairs;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(c) 10 pairs;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(d) 5 pairs.</td>
<td></td>
</tr>
</tbody>
</table>

Cognitive abilities assessment

The subjects’ cognitive abilities were assessed via the second edition of the Intelligence Test for Junior High School (Chen, 2004), which is a commonly used in Taiwan, and thoroughly standardized test that suitably assesses the diversity of cognitive abilities in adolescents. The Intelligence Test for Junior High School-II consists of three inventories: verbal reasoning, mathematics reasoning, and figural reasoning, and each of these inventories contains 39, 32, and 32 items respectively. The internal consistency of the Intelligence Test for Junior High School-II range from 0.79 ~ 0.88.

Genetic Foundation Test (GFT)

The GFT was utilized to assess students’ prior knowledge about genetics concepts before presenting the curriculum. The GFT consisted of 15 multiple-choice questions. These questions covered cell biology and basic genetic concepts that they had been taught in school. The test scores ranged from zero to 15. The content validity of the GFT was verified by two university professors of genetics and two experienced biology high school teachers. The internal reliability was calculated at 0.69 (Cronbach’s alpha) in the current study.

Cognitive Load Questionnaire (CLQ)

The extent of the cognitive load that the instructions presented was evaluated via subjective measures in the current study. Using a post-treatment questionnaire, participants were asked to report the amount of mental effort they devoted to learning the genetics materials. This strategy has been frequently utilized in cognitive load studies and found to be highly reliable (Paas, 1992; Paas & Van Merrienboer, 1994; Paas, Van Merrienboer, & Adam, 1994). Chang and Yang (2010) also suggested that the subjective assessment of cognitive loads is convenient for administration; thus, it is appropriate for use in science classrooms.

According to CLT (Sweller et al., 1998), intrinsic cognitive load refers to the inherent nature of the materials to be learned, which is fixed and innate to the instruction; extraneous cognitive load results from the instructional design itself, which may vary between different methods of presenting the learning content. The overall load reflects the experienced load based on the entire working procedure. Because the difference between animation and static picture instruction is the presentation format, if animation does help to reduce the cognitive load, it is
reasonable to anticipate that participants in the animation condition would report a lower level of extraneous load. The CLQ has 13 items, including the extraneous and overall cognitive load of the curriculum. To evaluate the extent of the extraneous cognitive load, students had to indicate how much mental effort they perceived related to the multimedia instructions. The extraneous cognitive load may be attributed to the design of the instructional materials, such as the manner in which the curriculum was presented and the user-interface operation (as shown in Table 1). To evaluate the extent of the overall cognitive load, students had to indicate the amount of the mental effort required for learning (e.g., “It required substantial mental effort to understand the course content relating to mitosis”). Each item was rated on a scale ranging from 1 to 7 (1: strongly disagree; 7: strongly agree). A higher CLQ score indicated that a greater cognitive load was perceived by students. The content validity of the CLQ was verified by two university professors in the science education research field. The internal reliability in this study was shown to be adequate, with the Cronbach’s alpha calculated to be 0.94.

The Genetics Concepts Test (GCT)

To assess students’ achievement on genetics learning, the GCT was administered after the curriculum. Table 1 includes examples of questions from the GCT. The GCT consisted of two sections that evaluated different aspects of achievement. The first section had three open-ended questions designed to assess students’ understanding of their genetics learning. Two experienced biology teachers were asked to grade students’ answers following the standard answers and scoring criteria provided. The second section contained 20 multiple-choice questions that aimed to evaluate students’ knowledge of the genetics they learned in the curriculum. Students received one point if they correctly answered each question in this section. The content validity of the GCT was verified by two university professors and two junior high school teachers in the field. The internal reliability (Cronbach’s alpha) was calculated to be .81 in the present study.

Data analysis

Since the assumption of normal distribution was not fulfilled for all dependent variables, the Mann–Whitney U-tests were conducted to evaluate differences in prior knowledge, perceived cognitive load, and learning outcomes between the animation and static pictures groups. The statistical tests were performed with SPSS version 18.0.

Results

Perceived cognitive load

The results revealed that the animation instruction required less cognitive load than the static picture instruction. As shown in Table 2, one of the major findings in this study was that students in the animation group perceived significantly lower extraneous cognitive loads than students in the static pictures group ($U = 3358.5, p = .037$). However, perceived overall cognitive load of the animation group was not significantly lower than the static pictures group ($U = 3592.00, p = .151$).

Learning outcomes

As the data in Table 2 shows, the animation group outperformed the static pictures group on the open-ended questions ($U = 2985.00, p = .002$). However, a statistically significant difference was found between the two groups on the multiple-choice questions ($U = 3795.00, p = .395$). It is worth noting that the animation group did not significantly differ from the static pictures group in terms of the GFT scores ($U = 4046.50, p = .892$), thus we can believe that the difference found in the open-ended questions was not due to different levels of prior knowledge.
Table 2. Comparisons of students’ perceived cognitive load in learning, prior knowledge, and genetic conception learning outcomes between the animation and static pictures group

<table>
<thead>
<tr>
<th>Cognitive Load Questionnaire (CLQ)</th>
<th>Mean (SD)</th>
<th>U</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Extraneous cognitive load</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Animation Group (n = 92)</td>
<td>3.66 (1.13)</td>
<td>3358.80</td>
<td>.037*</td>
</tr>
<tr>
<td>Static Pictures Group (n = 89)</td>
<td>4.11 (1.35)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Overall cognitive load</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Animation Group (n = 92)</td>
<td>4.37 (1.62)</td>
<td>3592.00</td>
<td>.151</td>
</tr>
<tr>
<td>Static Pictures Group (n = 89)</td>
<td>4.72 (1.61)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Genetic Foundation Test (GFT)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Animation Group (n = 92)</td>
<td>11.28 (3.05)</td>
<td>4046.50</td>
<td>.892</td>
</tr>
<tr>
<td>Static Pictures Group (n = 89)</td>
<td>11.43 (2.79)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Genetic Concept Test (GCT)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Open-ended</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Animation Group (n = 92)</td>
<td>15.28 (5.28)</td>
<td>2985.00</td>
<td>.002**</td>
</tr>
<tr>
<td>Static Pictures Group (n = 89)</td>
<td>13.46 (4.50)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple-choice</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Animation Group (n = 92)</td>
<td>10.83 (4.41)</td>
<td>3795.00</td>
<td>.395</td>
</tr>
<tr>
<td>Static Pictures Group (n = 89)</td>
<td>11.44 (4.23)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

Discussion

The present study aimed to compare the relative effectiveness of animation- and static picture-based multimedia instruction in invisible infinitesimal phenomena (i.e., genetics). Previous studies that compared the relative effectiveness of different multimedia platforms did not focus on the design equivalence of the multimedia platforms and the application of learning theory, which makes it hard to interpret the results. To refine previous studies, we attempted to design the animation and static picture instruction to be equivalent based on the principles suggested by CTL and CTML.

The first major finding of this study was that the animation group outperformed the static pictures group in the open-ended questions. A number of studies have indicated that multiple-choice formats may be appropriate for questions that assess the memorization of key points, facts, dates, and definitions. In contrast, open-ended questions that elicit students’ constructed responses and give students higher degrees of freedom in reasoning may serve as a better foundation for the evaluation of students’ higher-order reasoning and qualitative understanding (Chang, Yeh, & Barufaldi, 2010; Wang, Chang, & Li, 2008). Previous studies also showed that it is difficult to develop multiple-choice test items that assess higher cognitive skills or conceptual structures (Chang et al., 2010; Wang et al., 2008). The better performance of the students in the animation group on the open-ended question may be attributed to the finding that animation helps learners visualize the invisible infinitesimal process of genetics. This finding is in accord with viewpoints provided by several researchers. For example, Marbach-Ad, Rotbain, and Stavy (2008) suggested that animation can be used more effectively than static pictures to illustrate difficult abstract concepts and to visualize dynamic processes. Similarly, Russell, Netherwood and Robinson (2004) also suggested that animation in multimedia for the teaching of biology helps students to integrate and understand abstract concepts.

This study also provided evidence that students in the animation group perceived a lower extraneous cognitive load than students in the static picture group. As mentioned previously, animation is often considered too complex or too fast to be accurately perceived (Tversky et al., 2002). In other words, learners may perceive a larger cognitive load in animation-based learning. However, recent researchers in the area of cognitive psychology have reported a fascinating insight: the cognitive load imposed by instructional animation can be ameliorated by an appropriate instructional design (Hasler, Kersten, & Sweller, 2007; Mayer & Moreno, 2003; Moreno & Mayer, 1999). Our results provided empirical support that a well-designed animation instruction could generate a smaller extraneous cognitive load than a static pictures instruction. Since the two instructions were designed to be equivalent, except that the animation was dynamic, the difference could not come from the non-equivalence of the design. A possible explanation for why students in the animation group perceived a lower extraneous cognitive load than those in the static pictures group in the current study is that the transitional nature of the animation can guide learners’ attention, helping them focus on important information in the instructions. We are hopeful that future research will examine this possibility.
One should note that, we cannot conclude that lower perceptions of extraneous cognitive load gave rise to better learning performances in this study. We did not manipulate perceived extraneous cognitive load to investigate whether it consequently brings changes in students’ learning performances, the causal relationship between them therefore was not established. Although according to our data, after controlling the influence of prior knowledge, lower perceptions of extraneous cognitive load did predict higher scores of open-end questions in the GCT ($\beta = -0.133$, $p = .049$), which implies that these two variables were associated with each other, future studies are still needed to confirm the causal relationship.

Although this study supports the superiority of animation instruction, the findings should not be interpreted to mean that animations are relatively more effective in all situations. Since the instructions were designed to visualize processes that are not visible by the naked eye in the real word in this study, it is not appropriate to generalize the findings to instructions that are designed to visualize visible phenomena. Future studies are needed to examine the relative effectiveness of animations that are designed to visualize other types of phenomena compared to static picture instructions.

Moreover, this study should not be taken to controvert the value of the static pictures approach to aid students’ learning. Instead, static picture instructions might also be useful to support students’ learning in some aspects. Our findings indicated that the static picture instruction was not less effective for students’ performance of multiple-choice questions, which indicates that the static pictures approach could be an alternative and convenient way to help students acquire basic concepts of a learning topic.

An interesting and valuable topic for future research into multimedia learning in genetics would consider individuals’ working memory capacity (WMC). WMC is a frequently used index of an individual’s cognitive capacity, which is known to play a crucial role in higher-order cognitive functions. It has been suggested that the impact of cognitive load on individuals with lower WMC exceeds the impact on individuals with higher WMC (De Neys, 2006). Thus, it is reasonable to expect that a well-designed animation instruction would benefit individuals with lower WMC more than those with higher WMC. A number of studies have examined the effect of WMC on multimedia learning (Austin, 2009; Lusk, et al., 2009; Seufert, Schütze, & Brünken, 2009); however, the results are inconsistent. Future studies examining the effect of animation instruction on WMC would contribute to a better understanding of the boundary conditions of the multimedia approach.

In conclusion, this study, with consideration of equivalent design between instructions and the application of learning theories, reveals that the animation approach is more effective in helping students to learn invisible infinitesimal phenomena than the static pictures approach because it helps to lower the perceived extraneous cognitive load and reach higher achievement. Although this study is not without restrictions, it provides a methodologically sound comparison between the animation and static pictures approaches, which could serve as a basis for future studies.

Acknowledgments

The work in this study was supported by the National Science Council of Taiwan under contracts NSC 105-2511-S-008 -002 -MY3 and 105-2511-S-003 -012 -MY3 and by the Aim for the Top University Project-NTNU.

References


Effects of Multimodal Learning Analytics with Concept Maps on College Students’ Vocabulary and Reading Performance

Shih-Ping Wang1* and Yih-Lan Chen2

1Department of Applied Foreign Languages, National Taiwan University of S&T, Taipei, Taiwan // 2Department of Applied English, Ming Chuan University, Taoyuan, Taiwan // spwang2005@yahoo.com.tw // ylc1958@gmail.com

†Corresponding author

(Submitted December 5, 2016; Revised April 30, 2017; Accepted June 2, 2017)

ABSTRACT
This study integrated the multimodal framework of learning analytics (IMFLA) with the concept mapping (Cmap) approach to improve students’ vocabulary and reading abilities. A total of 70 participants were divided into 2 classes, Class 1 (experimental) and Class 2 (control), for a 1-year period. Vocabulary and reading tests were implemented 3 times. Repeated measures were conducted to test the effect of the program. The results indicated that Time had a significant effect on enhancing both outcome measures (p < .01 for vocabulary; p < .001 for reading). Moreover, significant interaction effects between Time and the program on both vocabulary (p < .05) and reading (p < .001) further suggest that a longer period of time spent on the program would result in a significant effect on both the vocabulary and reading outcome measures. That is, a significant interaction effect occurred between IMFLA and the time factor. Such interaction effects resulted in better vocabulary and reading abilities when students of Class 1 spent more time on the IMFLA procedure. The multimodal and learning analyses of students’ weekly logs confirmed this improvement. We therefore suggest that instructors use digitalized wordlists and Cmaps in language instruction to enhance students’ vocabulary and reading abilities.

Keywords
Multimodal, Learning analytics, Concept map, IMFLA

Introduction
It has long been recognized that vocabulary knowledge is closely related to reading comprehension (Qian, 2002). Vocabulary learning through reading has direct influences on classroom practice (Nation & Webb, 2011). On the one hand, language instructors often use concept mapping (Cmap) in reading activities to enhance reading comprehension (Wang, Huang, & Hwang, 2016). Learning analytical methods are further needed to observe the details of learning processes, students’ language use and learning behavior.

On the other hand, the process of learning to read through digital approaches is rather complicated. Such a process involves three critical components (Park, Zheng, Lawrence, & Warschauer, 2014): word recognition (e.g., reading development and software for word recognition), language comprehension (e.g., multimedia glosses, vocabulary learning, and corpus-based wordlists), and text interpretation (e.g., textual structures, digital reading materials, translation, and interactive feedback). However, few studies have investigated the process of applying Cmap to vocabulary and reading learning by making a book using digital methods. To probe such a complicated process, the current study aimed to explore the effect of integrating multimodal learning analytics (LA) with Cmap on students’ vocabulary and reading abilities.

First, as the popularity of computer technology has brought multimodal approaches into the digitalized epoch, many studies have demonstrated the effectiveness of Cmap in language and content learning (Liu, Chen, & Chang, 2010). The computerized approach has simplified the Cmap learning process, which enables more flexible presentations and interactions with the learning content (Hwang, Wu, & Ke, 2011; Yang, Hwang, Hung, & Tseng, 2013). It is evident that students understand better and remember more while learning from a Cmap than from a textual presentation (Kim & Olaciregui, 2008). In addition, Cmap is considered a technical tool for LA studies; it is used for students to organize and visualize knowledge experience, which can be tracked to support LA studies in ubiquitous environments (Hwang, Hung, Chen, & Liu, 2014).

Second, LA has been developing rapidly in recent years. It uses intelligent data, learner-produced data, and analysis models to discover information along with social connections to predict and advise on learning (Siemens, 2010). Johnson, Adams, and Cummins (2014) defined LA as an educational application of “big data” within online and hybrid environments.
Literature review

Corpus-based wordlists and reading comprehension

A frequency list has long been considered an essential tool of corpus linguistics (Barker, 2010). Frequency, word lists, and the corpus should be considered in vocabulary teaching (Nation & Webb, 2011). The benefits of using corpora for creating word lists for vocabulary learning are well recognized. There are two main considerations regarding wordlists for L2 learners (Coxhead, 2011). First, frequency plays a significant role in creating wordlists. Second, digitalized or online L2 vocabulary learning is closely related to reading ability (Abraham, 2008; Cobb, 2007). Tribble and Jones (1997) argued that the most effective starting point to appreciate a text is a frequency-based word list. For instance, a substantial frequency list can be compiled and derived from the British National Corpus (BNC), a large corpus of modern English (Burnard, 2007). Additionally, one of the most commonly used wordlists is the Academic Word List (AWL), selected from the Academic Corpus (Coxhead, 2011). However, few studies have explored how a learners’ corpus used together with Cmap can be analysed, or how Cmap can be used in teaching vocabulary to reinforce reading ability.

Concept map analysis

A Cmap is considered a technical tool for LA studies (Hwang, Hung, Chen, & Liu, 2014). According to Hughes, Maccini, and Gagnon (2003), visual displays can be shown in temporal, sequential, hierarchical, semantic and comparative patterns. Each pattern has its particular function and purpose to arrange the information explicitly. For example, graphic organizers are based on the way they arrange information, typically including four general types: hierarchical, conceptual, sequential and cyclical patterns (Novak & Cañas, 2007).

Concept maps are also used in facilitating cooperative learning (Novak, 1991). Especially, many studies have shown the effectiveness of Cmap for engaging students in meaningful learning. For example, Hwang, Hung, Chen, and Liu (2014) reported that Cmap is an assessment tool for helping instructors evaluate students’ cognitive levels and knowledge structures. In their “Mindtool- Assisted In-field Learning” project, Cmap was employed in context-aware ubiquitous learning activities, and the authors claimed that the students’ learning achievements and attitudes were significantly improved.

Cmap can also improve students’ literal and inferential comprehension ability. Dias (2010) suggests that the application of the Cmap strategy may enhance L2 learners’ reading comprehension. However, the online learning processes of Cmap should be observed, and the learners’ behavior and language use should be noticed and analyzed through learning analytics to improve students’ vocabulary and reading proficiency.

Learning analytics

Learning analytics (LA) is the use of data produced by the learner, and analysis models to discover information and social connections in order to offer advice for learning (Siemens, 2010). More recently, the 2016 Horizon Report (p. 38) highlighted LA as “an educational application of web analytics aimed at learner profiling, a process of gathering and analyzing details of individual student interactions in online learning activities.”

The data analysis of LA has developed rapidly and requires both quantitative and qualitative analyses to improve the student learning experience. LA plays an important role in learning processes through online learning logs, insights, and information for instructors and learners (Hwang, Hung, Chen, & Liu, 2014). In other words, LA includes the following features (Johnson, Adams, & Cummins, 2014):

- Using “big data” and adaptive learning data,
Using learner-produced data to discover information,
Individualizing personal learning experience,
Providing online-interactive feedback,
Transforming education “from a standard one-size-fits-all delivery system” into a flexible framework to meet students’ needs and interests.

The multimodal approach

Multimodal methods deal with communicative language in different modes and other devices. Herring (2015) proposed a framework of multimodal computer-mediated communication (CMC), including (1) interactive multimodal platforms—used to support a union of modes such as text, audio, video, and graphics for user-to-user communication; and (2) robot-mediated communication—human-human communication through “voice, video, and motion” in physical space via a remotely controlled robot.

In other words, language is not the only approach and source for conveying meaning and informing one’s social identities (Cameron & Panovic, 2014). Language must cooperate with other communication modes, which refers to other non-verbal signals, such as facial expressions, gazes, gestures, and the like, to achieve the characteristics of language through which people can express the meaning of their utterances and identities. Sindoni (2014) claimed that there are three ways in which we gain information from non-verbal resources and language cues: (1) the speaker’s facial expression, (2) kinetic action, and (3) proxemics, which can be used to analyze the classroom discourse. “Kinetic action” is related to body movement such as head-nodding and hand-waving, while “proxemics” refers to the distance between speakers.

One specific approach to delving into how language is integrated with other communication modes is multimodal discourse analysis. Such analysis intends to investigate how other tools and other multiple modes (e.g., gaze, gesture, and proxemics) are combined together with discourse (Jones, 2012). One key principle of multimodal analysis is that the aim is not to understand how those outside factors exploit their power in language, but rather, to explore how language “works together” with other multiple communication modes.

Research questions

The current study integrated a multimodal framework of learning analytics (IMFLA). We investigated its effect on all the participants’ vocabulary and reading abilities during a period of two academic semesters. Accordingly, the following research questions were proposed to guide the current study:
RQ 1: Is Time effective in terms of improving students’ (1a) vocabulary and (1b) reading test performance?
RQ 2: Is IMFLA effective in terms of improving the students’ (2a) vocabulary and (2b) reading test performance?
RQ 3: Is there any interaction effect between Time and IMFLA on the students’ (3a) vocabulary and (3b) reading test performance?
RQ 4: How did the learning analytics support the log analysis, concept mapping and multimodal approaches?

Method

Participants

A total of 70 students from a public university in Taiwan in two intact classes: Class 1 (experimental group; n = 35) and Class 2 (control group; n = 35) participated in this study. They were students enrolled in a Vocabulary and Reading course. The students’ English vocabulary and reading abilities were evaluated three times, first at the beginning of the course, second time at the end of the first semester, and finally at the end of the second semester.

Materials, treatment, and course design

The vocabulary teaching material, a textbook entitled Most Essential Prefixes and Word Roots (Hsu, 2006), was adopted to improve the students’ advanced academic vocabulary. Students were provided with a list of the first 900 words extracted from this textbook based on their frequency ranking in the BNC. Besides the extracted wordlist, the online Cmap was also used for vocabulary teaching.
We tried to use concept maps to examine students’ word power and engaged them in the collaborative construction of Cmaps. *Inspiration 9* was used to draw Cmaps. Only students of Class 1 were encouraged to use any Cmap tools to draw Cmaps. For example, Figure 1 demonstrates examples of two types of Cmap. Based on two domains (*prefix* and *root*) and propositions (= linking lines with linking words), a prefix “ad-” (“to”) linking with a root “-orn” (“to deck”) becomes “adorn” (“to decorate”). Likewise, the prefix “contra-” (“against”) linking with a root “-dict” (“say”) becomes “contradict” (“to oppose”).

![Cmap for teaching prefix ad- (Web-like pattern)](image1)

![Cmap for teaching prefix contra- (Sequential pattern)](image2)

*Figure 1. Cmaps of vocabulary teaching and learning*

As for the reading materials, both classes were required to read English news articles and short essays. They watched the movies *Pride and Prejudice* and *Dead Poets Society*, read the novels on which the movies were based, and then translated one of them into Chinese as their final project, i.e., their own translated book. There were six small teams (5-6 students per team) in each class; but, the teaching methods were different. In Class 1, the multimodal method was made available to the students with the Cmap practices and the electronic, interactive, or online materials to explain the plot, characters, facial expression, body language, and gestures of the characters in the novel. In Class 2, the traditional grammar-translation approach was mainly used without interactive Cmap exercises or the electronic online materials.

All of the students took three reading and vocabulary tests, but only Class 1 maintained online weekly logs used as a corpus for further learning analysis. Logs for 36 weeks were collected, 18 for each semester. To observe the learning behavior, each student’s weekly log was analyzed, including three sections for the analysis of word frequency, wordlist, and concordances: (1) What I have learned today, (2) How I helped my classmates, (3) Suggestions for the instructor, who could then reply to the student’s feedback or give online comments immediately. Based on these criteria, the logs entered by the students in Class 1 were accessible to the in-group participants and the teacher. Table 1 summarizes the course design, including the vocabulary and reading materials.

<table>
<thead>
<tr>
<th>Table 1. Course design</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Item</strong></td>
</tr>
<tr>
<td><strong>I. Vocabulary</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>II. Reading</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>III. Outcome</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Measurement

Quantitatively, the students’ performance was measured via the vocabulary and reading tests. The vocabulary tests (Wang, 2006) were designed based on a modified academic wordlist with example sentences from the BNC (see the sample test in Table 2).

A pilot test (using test-retest) was administered (n = 345; average= 65; higher group = 84, lower group = 44) for the reliability coefficient (the test-retest coefficient = .74, p < .01; Cronbach’s alpha = .97) and the Predictive criterion-related validity was evaluated using the vocabulary test and Joint College Entrance Exam (the predictive validity coefficient = .76, p < .01). The reading test was based on the TOEFL reading test (three short essays with 10 multiple choice questions). Since each dependent measure was implemented three times, test-retest reliabilities were calculated. The results indicated that the reading and vocabulary tests reached acceptable reliability, 0.73 and 0.77, respectively.

Qualitatively, students of Class 1 maintained weekly online logs collected as a corpus (included in IMFLA), which is the difference between these two groups. A summary of the material used for testing and outcome measures was the wordlist, modified AWL for testing only, the TOEFL Reading test, the Project (a book), and the Weekly logs (Online interactive for Class 1).

Table 2. Item analysis for example questions (50 questions)

<table>
<thead>
<tr>
<th>Vocabulary questions with example sentences</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
</table>
| 1. ( ) Search facilities as well as browse facilities are available now.  
  a. useful  b. reliable  c. obtainable  d. valuable | .000*** |
| 2. ( ) This formula, very common in law examinations, means “Can B sue A successfully?”  
  a. form  b. rule  c. movement  d. fortunate | .005** |
| 3. ( ) As we have seen they may involve other types of expert as well.  
  a. include  b. invent  c. resolve  d. solve | .010 |
| 4. ( ) It was this principle that seemed now to be inadequate.  
  a. principal  b. prime  c. answer  d. law | .016* |
| 5. ( ) I am sure that our readers have their pens ready to respond.  
  a. respect  b. resign  c. resolve  d. reply | .003** |

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

Analysis

Figure 2 is the proposed model (i.e., IMFLA) involving four equally important modes (adapted from Herring, 2015): The stakeholders (instructor-learners), data (weekly logs → corpus), audio-video (clips), and instruments (Cmap, WordSmith Tools, etc.) in addition to an interactive multimodal platform based on the online Blackboard to maintain the consistency of the different modes.

Multimodal Framework for Learning Analytics

Figure 2. Integrated multimodal framework for learning analytics (IMFLA)

In this study we adopted repeated measures to answer research questions 1-3, using SPSS 22 statistical software. The current design measured the effect of IMFLA over three points of time, treating reading and vocabulary abilities as dependent measures, while IMFLA was a major factor. There are two reasons for the choice of repeated measure: (1) to enhance the power to detect effects, and (2) to detect interaction effect.

According to Seltman (2018) and Loerts (2008), when the same variable is measured more than once for each subject, the use of repeated measure reduces unsystematic variability in the design, resulting in greater power to
detect effects. Furthermore, Seltman (2018) suggested that the primary purpose of repeated measure is to detect an interaction effect. If the interaction is significant, then both factors, the IMFLA and time, affect the outcome. To answer research question 4, WordSmith Tools was used to analyze the online weekly logs in terms of word frequency (%), wordlist, clusters, and concordance.

Procedure

All participants took reading and vocabulary tests before the launch of the course. An online Blackboard was used as a platform for IMFLA. The procedure is summarized in Figure 3.

![Figure 3. The IMFLA procedure](image)

After the first test, the Class 1 students learned English vocabulary using the corpus-based wordlist with Cmap-interactive word parts and reading, and then wrote their online logs (36 weeks). They also translated the novel and submitted it as a book (learner-produced data). Additionally, their online weekly logs were collected and downloaded as an electronic corpus for learning analysis to discover information, to predict and advise on learning in order to improve their learning. They took the second set of tests at the end of the first semester and the final set of tests at the end of the second semester.

However, students of Class 2 used a relatively more traditional approach. They memorized English vocabulary focusing on word parts (i.e., prefixes, roots, and suffixes), wrote assignments, and submitted hardcopies of their assignments. Both classes finished translating a movie novel and submitted it as a book by the end of the second semester. They also took the second tests and the final tests.

Results

Descriptive statistics

Table 3 shows the results of the descriptive statistics. The students’ vocabulary scores increased from 83.89 to 86.74 and further to 88.23 for Class 1, while they increased from 88.69 to 89.20 but then decreased slightly from V2 to V3 (89.20 to 89.14) for Class 2. As for the students’ reading scores, they increased from 57.43 to 74.29 and further to 79.71 for Class 1, while they increased from 80 to 90 but then decreased to 85.43 for Class 2.
Tests of within-subject and between-subject effects for vocabulary ability (RQ1a-3a)

To investigate the differences in the vocabulary scores in repeated measure, Mauchly’s W was first implemented for the Sphericity test. The result ($p = .411 > .05$) indicated that the variances in the differences between all possible pairs of levels of vocabulary tests were equal; that is, there was no violation of repeated measure assumption; therefore, Sphericity could be assumed. As a result, a test of within-subject effect (repeated measure) was considered. To answer RQ1a, Repeated measure for within-subject effect on vocabulary scores indicates that Time was significant in enhancing the learners’ vocabulary ability ($F = 5.460, p < .01$, partial $\eta^2 = .074$), as shown in Table 4.

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III SS</th>
<th>df</th>
<th>Mean square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial eta squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tests of Within-Subject Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>212.610</td>
<td>2</td>
<td>106.305</td>
<td>5.460</td>
<td>.005</td>
<td>.074</td>
</tr>
<tr>
<td>Time * Class</td>
<td>133.981</td>
<td>2</td>
<td>66.990</td>
<td>3.440</td>
<td>.035</td>
<td>.048</td>
</tr>
<tr>
<td>Tests of between-Subjects Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class</td>
<td>389.505</td>
<td>1</td>
<td>389.505</td>
<td>1.926</td>
<td>.170</td>
<td>.028</td>
</tr>
</tbody>
</table>

As for the answer to RQ2a (vocabulary), the between-subject analysis (see Selman, 2014) indicated that IMFLA’s effect on the learners’ vocabulary ability was insignificant ($F = 1.926, p = .170 > .05$), with only 2.8% of the variance in vocabulary ability accounted for by IMFLA. The analysis also confirmed a significant interaction effect between Time and Class on vocabulary ability ($F = 3.44, p < .05$, partial $\eta^2 = .048$), and thus answered RQ3a (vocabulary). The two figures of partial $\eta^2$ (effect size) suggest that 7.4% of the variance in vocabulary scores was explained by the Time factor, while 4.8% of the variance in reading scores was explained by the interaction between Time and the treatment.

Results of within-subject and between-subject effects for reading ability (RQ1b-3b)

Mauchly’s W was conducted to test for Sphericity. Table 5 showed that the variances in the differences between all possible pairs of levels of reading tests were equal ($p = .093 > .05$). With no violation of the repeated measure assumption, Sphericity could be assumed. As a result, a test of within-subject effect (repeated measure) was considered.

To answer RQ1b (reading), repeated measure shows that Time was significant in enhancing the learners’ reading proficiency ($F = 28.157, p < .001$, partial $\eta^2 = .293$). Again, to test whether IMFLA affected the learner’s reading ability in two groups (RQ2b), a between-subject test was implemented. Results indicated that the effect of IMFLA was significant ($F = 23.924, p < .001$, partial $\eta^2 = .260$). The effect size, partial eta squared of .260 indicates that 26% of the variance of reading proficiency can be accounted for by IMFLA.

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III SS</th>
<th>df</th>
<th>Mean square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial eta squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tests of Within-Subject Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>8692.381</td>
<td>2</td>
<td>4346.190</td>
<td>28.157</td>
<td>.000</td>
<td>.293</td>
</tr>
<tr>
<td>Time * Class</td>
<td>2515.238</td>
<td>2</td>
<td>1257.619</td>
<td>8.148</td>
<td>.000</td>
<td>.107</td>
</tr>
<tr>
<td>Tests of between-Subjects Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class</td>
<td>11293.333</td>
<td>1</td>
<td>11293.333</td>
<td>23.924</td>
<td>.000</td>
<td>.260</td>
</tr>
</tbody>
</table>
The analysis also confirmed that there was a significant interaction effect between Time and Class on reading ability \( (F = 8.148, p < .001, \eta^2 = .107) \), as a positive answer to RQ3b (reading). In other words, with both significant effects, the two figures of partial \( \eta^2 \), indicating effect size, suggest that 29.3% of the variance in the reading scores was accounted for by the Time factor and that 10.7% was accounted for by the interaction between Time and the treatment.

**Vocabulary (V) and Reading (R) tests at three points in time**

Visual presentation of the interaction effect is illustrated in Figure 4 where Class 1 kept improving their reading ability even after the experiment at the end of the second semester, while Class 2 improved through whatever approach but showed no further improvement after one more semester.

![Tests at three points in time](image)

*Figure 4. Vocabulary (V) and Reading (R) tests at three points in time*

This was also true for the smaller effect size of IMFLA on vocabulary \( (\eta^2 = .074) \) as the result of the ceiling effect connected to the relatively high beginning level of the vocabulary tests. Additionally, our pilot study indicates that the average score of the higher group is 84, and Class 1 continued increasing their scores (83.89 → 88.23) significantly, which “broke the bottleneck of the ceiling effect.” That is, the IMFLA helped students keep improving their vocabulary and reading abilities through the interaction effect.

**Learning analytics results**

**Statistical results and textual analysis for weekly logs**

To answer RQ4, the following statistical results and textual analysis are presented in detail. Weekly logs were constructed as a learners’ corpus (=172,607 running words) and downloaded from the online interactive Blackboard. For example, Figure 5 shows that the top four words (i.e., feedback, reflection, suggestion, and comment) are related to students’ learning experience. The “concept map” (frequency = 216) refers to what (contents), how, and how often (behavior) students learned and used it in vocabulary and reading. Both involve the Cmap in the students’ logs with relevant lexical items (e.g., words, word, vocabularies, and vocabulary).

Figure 6 highlights the concordance of “concept map” (frequency = 216; dispersion value = 0.63, displaying how “concept map” is evenly distributed in the text/corpus), revealing the students’ learning behavior or learning contents. The frequently used 3-word clusters such as “the concept map” (freq. = 86), “a concept map” (43), and “draw a concept” (freq. = 21) demonstrate how and how often students used this Cmap skill in their learning. It seems that they appreciated the Cmap method to learn both reading and vocabulary, which was supported by their feedback. The weekly log information could help the instructor adjust the teaching tempo, encourage students, or solve individual learning problems immediately.
Table 6 indicates how an individual learner (student-1) interacted with the instructor (stakeholders), and how she benefited from the Cmap method in terms of her vocabulary and reading learning with her teammates. She expressed how she enjoyed using Cmap for learning vocabulary and reading through online interactions: for example, “The concept map will really be useful,” “It is a good tool to help me learn,” and “helping us understand the idea of concept map.” Finally, the instructor answered her and adjusted the teaching pace or content to meet the learner’s needs. Student-2 also expressed that she used Cmap to improve her vocabulary and reading ability. It was predictable that the students enjoyed the Cmap methods, and performed better later on. Table 7 indicates how a student drew a concept map based on her personal learning experience. It also shows her learning behavior and her interaction with the instructor.
Table 6. Student’s learning log and comments on Cmap

■ Student-1:

(1) What I have learned today
I experienced the importance of teamwork again. We got two handouts of vocabularies, which should be scheduled and finished by ourselves. We should work with our team members. We had a discussion about how to finish those vocabularies in the class. Also, I learned some general ideas of concept map. It has four types of diagram which can help us understand the system or relationships among ideas. I think concept map will really be useful in our daily lives.

(2) How I helped my classmates
2.1 You told us the application of concept map. It is a good tool to help me learn.
2.2 We had a quiz. I tried my best to understand the organization of the article. After the quiz, you gave us two vocabulary handouts. Our group shared ideas of how to schedule our time on studying these vocabularies. I thought that we all had a good time in these two classes.
2.3 We shared our ideas of study plan with each other. We found a way to share the work of finding the meaning of vocabularies. Then, we had a small activity of which purposes were helping us understand the idea of concept map. We shared our thought of which word we should put in the blank.

(3) Suggestions for the instructor
I got some information from the class. You let us know the website BNC. You told us the four types of concept map. You told us that we should memorize vocabularies by synonyms or English sentences. I hoped you to share us more and I expected what you would share with us next class.

■ Instructor: Thank you for your comments. I am glad you like concept mapping method. Yes, I will share more about the concept map.

■ Student-2: We will use concept map to improve our vocabulary and reading ability…

Table 7. Weekly log example about Cmap

(1) What I have learned today
You guided us to find out the topic sentence step by step. You taught us to observe those words with repetition and cohesion so that we could find out their relation more easily. I really love this skill...
Furthermore, you asked us to use concept map. After drawing out a concept map, I could understand their relationship mentioned in the article. I learned the importance of the concept map again!

Statistics, multimodal analysis and the products

Figure 7 displays the multimodal and multimedia terms with their frequencies (e.g., movie, movies, videos, film; movie titles such as DPS and P&P). These words were frequently used in the weekly logs. It seemed that watching the movie was popular among students.

As shown in Table 8, it became easier to understand the plot with the help of the Cmap method through the learning analytics. Table 8 indicates that the students enjoyed the novels. For example, Cindy considered Cmap as useful for organizing ideas, while Shanna “was fascinated by the plot” with the help of Cmap. Kathy was also drawn to the synopsis. Likewise, Hanna looked forward to watching the remaining portion of the movie.
Table 8. Comments of four students on the movies

<table>
<thead>
<tr>
<th>Student</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cindy</td>
<td>I learned some <strong>useful skills</strong> to organize my ideas such as the <strong>concept map</strong>... I am glad to know that <strong>you do take our comments seriously</strong> because that means we can share what we really want to learn in this class and you would take it into consider.</td>
</tr>
<tr>
<td>Shanna</td>
<td>... A special thing we did was <strong>watching the movie</strong> <em>Dead Poets Society</em>... I was fascinated by the <strong>plot</strong>... With the help of <strong>concept maps</strong>, it is <strong>easier to grasp the ideas</strong> of every paragraph and the whole article.</td>
</tr>
<tr>
<td>Kathy</td>
<td>...Surprise, surprise, boy do I love surprises. Indeed Professor, you sure surprised me with a <strong>movie</strong> today! Can I ever be more grateful for your generosity? Thanks for giving us hints on what lines we should remember in the <strong>movie</strong>. Quite frankly, although we haven’t even gone through <strong>DPS</strong>, <strong>I am already drawn to the synopsis</strong>.</td>
</tr>
<tr>
<td>Hanna</td>
<td>...Today, we practiced writing the <strong>concept map</strong> of Unit five... Later, we watched the first part of the movie <em>“Dead Poet Society,”</em> ...Today, we drew the <strong>concept map</strong> of Unit five together, which made our ability of outlining an article better. ...I really look forward to watching the <strong>movie</strong> next Wednesday!</td>
</tr>
</tbody>
</table>

The commonly used aspects for multimodal methods (Sindoni, 2014) include “gesture/facial expression,” “kinetic action” (body language), and “proxemics” (distance). Figure 8 presents the analysis and learning behaviors of our students, demonstrating how a student (Todd) played a shy boy in terms of these stages: (I) Facial expression, (II) Kinetic action, and (III) Proxemics. With the help of video clips, students understood better the text of the novel through these three stages and analyses.

![Figure 8. A clip of DPS using multimodal analysis](image-url)
Finally, Figure 9 indicates that the students worked in groups to make their own books (products) – a translation of the novel. The topic, contents, and supplementary materials of a translation novel were added and modified to some extent according to the course design. We used learner-producing data to discover new information and adjust our teaching contents.

(a) Translation of Dead Poets Society (b) Translation of Pride and Prejudice

*Figure 9. Six example books as the product*

**Discussion and conclusions**

In summary, both repeated measure analyses (for reading and vocabulary) indicated the significant interaction effect of IMFLA and time, suggesting that both IMFLA and time affected students’ reading and vocabulary abilities as suggested by Seltman (2018). Further analysis of the between-subject effect illustrating IMFLA has significantly affected learners’ reading ability. The effect size shows that 26% of the variance of reading ability is explained by IMFLA while only 2.8% of the vocabulary ability is accounted for by IMFLA due to a near ceiling effect.

On the one hand, students who scored near the ceiling on the pre-test had a relatively small opportunity for improvement in their scores on their post-test. Our vocabulary pilot test indicated that the average of the higher group was 84. In fact, the current study indicates that Class 1 kept improving its vocabulary ability (83.9 → 88.3). The repeated measure results demonstrated that IMFLA was effective in terms of enhancing students’ word power and reading ability. A significant interaction effect was also found between IMFLA and the time factor. This effect resulted in better vocabulary and reading comprehension when the students spent longer on the IMFLA procedure.

On the other hand, learning analytics and log analysis further confirmed that the reinforced application of multimodal approaches is an effective procedure for enhancing students’ vocabulary ability and reading proficiency. The results of the analyses are briefly summarized below:

- The weekly log in the online Blackboard as the platform is a proper channel and environment for language teachers and learners (stakeholders).
- The instructor could answer the individual questions, adjust the course contents, and reschedule the schedule immediately according to the online weekly logs.
- The quantitative and qualitative results in the logs indicated learning behaviors such as what (contents), how (learning behaviors), how often (frequency) students used or learned specific words and how they created their own Cmaps to learn or to draw the plot of an essay or a novel.
- Students’ learning behaviors were recorded through the online weekly log of the electronic platform, which helped the instructor revise his/her teaching contents.

This study has confirmed the advantage of IMFLA in the process of vocabulary and reading learning. First, we rejuvenated the textbook by using the frequency-based wordlist for the Class 1 students, and used Cmap as a tool to teach word parts. Second, the IMFLA approach was integrated into the instruction to enhance students’ reading comprehension. The weekly log analysis indicated that Class 1 appreciated the Cmap method for vocabulary and reading learning. It was found that the grammar-translation approach (Class 2) did not help the students to a significant extent, but the digitalized design (IMFLA) helped learners significantly increase their vocabulary and reading scores (Class 1).

Overall, the IMFLA demonstrated a significant effect on students’ vocabulary and reading abilities over time. Therefore, this study contributes to highlighting the importance of using Cmaps and frequency-based wordlists to teach word parts. Another contribution is that it demonstrates how the multimodal and Cmap approaches can
be integrated to enhance students’ reading. Finally, the texts of weekly logs as a corpus were analyzed to support the findings of our experimental study.

Although the current study showed that IMFLA can be beneficial to students’ vocabulary and reading abilities over time, training is required before teachers and students can make good use of the multimodal approach. Another limitation of the current study lies in the intact grouping of classes which is inherent in most educational settings. To counter this limitation, we used the repeated measure approach since there were measures at three points in time. However, to highlight the effect of the treatment, it would have been better to conduct the experiment for a longer period of time, with measures at more than three points in time.

Our proposed framework has the potential to become a widely-used learning model, and future studies should continue to investigate:

- how multimodal and learning analytic methods are integrated into IMFLA to apply the Cmap method, observe language learning behaviors, and give any feedback to improve teaching and learning quality.
- how video or novels based on movies are incorporated into IMFLA to assist students’ vocabulary and reading learning.

Finally, the development of effective multimodal vocabulary learning and digitalized reading tasks are also needed to improve pedagogical instruction.

Acknowledgements

This project was partially supported by the Ministry of Science and Technology, Taiwan, ROC (MOST: 104-2410-H-011-022-MY2).

References


Developing a Multidimensional Framework for Analyzing Student Comments in Wikis

Xiao Hu1*, Christy Weng-Lam Cheong2 and Samuel Kai-Wah Chu1
1Faculty of Education, University of Hong Kong, Hong Kong // 2Macao Polytechnic Institute, Macao // xiaoxhu@hku.hk // wlcheong@ipm.edu.mo // samchu@hku.hk
*Corresponding author

(Submitted June 13, 2016; Revised January 27, 2017; Accepted June 23, 2017)

ABSTRACT
This study develops a framework for analyzing student comments in Wikis of group writing to inform learning assessment. It first drew on the literature to develop a framework consisting of three modules measuring student interaction, meaning construction and thinking development in the writing process. In-service teachers were interviewed to ensure framework practicality and inform subsequent refinement. A sample of 1,482 Wiki page comments was collected from 48 groups of secondary school students in Hong Kong to test the developed framework. Statistical analyses and association rule mining were conducted to the coded data to explore the relations among coding categories. This study aims to raise the attention on page comments in the analysis of student activities in Wiki and provided empirical evidence on category relations, which will be instructive for further research and practice in Wiki-supported learning.

Keywords
Wiki, Student comments, Social interaction, Meaning construction, Thinking development

Introduction
Wiki, as a social media application, allows users to develop contents collaboratively. It is regarded as a useful tool to facilitate project-based learning activities (Li, Chu, & Ki, 2014; Lo, 2013; Wang, 2014). An increasing number of studies are devoted to explore its usage and affordance, demonstrating its values in strengthening student collaboration and facilitating knowledge acquisition (e.g., Aydin & Yıldız, 2014; Cullen, Kullman, & Wild, 2013).

A Wiki is made up of pages contributed by users. Each page consists of content and comments. The content part is where group writing is developed and presented, and for which revision history is tracked. The comment part is where individual users may leave short messages for their collaborators. According to Du, Chu and Chan (2016), comments on Wiki pages are closely related to various activities students perform in Wikis. Some comments facilitate communications by criticizing ideas; some point out writing issues, and others facilitate group coordination and collaborations (Du et al., 2016). When responding to each other’s comments, students may possibly engage in further discussions or page revisions, leading to an impact on the effectiveness or quality of collaborative writing (Judd, Kennedy, & Cropper, 2010; Woo, Chu, & Li, 2013). Figure 1 illustrates an example Wiki page with comments.

Wiki data are potentially useful for analyzing and monitoring students’ engagement and writing progress. The question is then how these data can be effectively analyzed to inform learning assessment. Existing studies on Wikis in education mainly focus on page content and edits (e.g., Macfadyen & Dawson, 2010; Romero-Zaldivar, Pardo, Burgos, & Kloos, 2012; Romero, López, Luna, & Ventura, 2013) while page comments are largely ignored. Analyses on page content and edits aimed to assist teachers in formative and summative assessment (e.g., McKenzie et al., 2013; Williams, 2014) and in identifying and monitoring students’ learning behaviors (e.g., Berland, Martin, Benton, Smith, & Davis, 2013; Brooks, Erickson, Greer, & Gutwin, 2014; Tobarra, Robles-Gómez, Ros, Hernández, & Caminero, 2014). However, few studies have investigated students’ comments made on Wiki pages. Students’ page comments in Wiki can provide additional evidences of student interactions and contributions. With proper processing, they can be made use by teachers in understanding, assessing and monitoring students’ learning. This study attempts to fill this research gap by developing a framework for analyzing students’ page comments in Wiki that can inform learning assessment. Specifically, the following research questions will be answered:
RQ1: How can student comments be categorized for facilitating teachers in learning assessment?
RQ2: Are there any relations among the categories of student comments?

To answer RQ1, a categorization framework was developed from the literature. In-service teachers were interviewed to inform modifications. The refined framework was then tried out on a sample of secondary school users...
students’ comments on their group project Wikis to inform further framework refinement. To answer RQ2, statistical analyses and association rule mining were conducted to the coded data, enabling discussions on the relationships between coding categories. This study helps raising the attention on page comments in the analysis of student activities in Wiki and providing empirical evidence on the theoretical and practical values of the proposed framework. This will be instructive for research and practice in Wiki-supported learning.

**E. Background of Study**

Our Focusing Question:
- 1. What's the meaning and definition of Online Exercise?
- 2. Why so many schools in Hong Kong use Online Exercise?
- 3. Are Online Exercise effective and improve our skills in that subject? Maybe it is just a waste of time?
- 4. Can Homework or work be replaced by Online Exercise?
- 5. Can Online Exercise prevent the problem of copying others answer?

Since we are a secondary school student, there are more online exercises that we need to do like English Builder, Pond of Reading (有簡天地). Also, it became more important to us because the result of your online exercises will show up on the report card. But are online exercises really that important and meaningful? We will check it out though this project study.

![Figure 1. An example Wiki page with comments](image)

**Literature review**

**Wiki as a tool for collaborative learning**

Collaboration, communication, critical thinking and problem solving are among the identified 21st century skills (Bruett, 2006; Partnership for 21st Century Skills, 2009). Wiki has been found to be a powerful platform that facilitates training and promoting these skills, for it can support an autonomous, collaborative and inquiry-oriented learning environment (Hakkarainen & Sintonen, 2002). In Wiki, students collaboratively write on the same pages, through communication and negotiation, among themselves and with teachers. This process leads to the co-construction of knowledge, as grounded in the social constructivist theory (O’loughlin, 1992). By allowing students to comment and modify each other’s writing in a cumulative fashion, Wiki motivates and helps students practice critical thinking (Cabiness, Donovan, & Green, 2013; Wake & Modla, 2012). In particular, through commenting on Wiki pages, students were found to have enhanced their social competences and metacognitive skills (Notari & Doebeli, 2012).

Current research on student assessment in the Wiki context seems to fall into two main streams. The first stream tends to suggest ways of integrating student peer assessment into the assessment framework (e.g., Šerbec, Strnad, & Rugelj, 2010; DeWever, van Keer, Schellens, & Valcke, 2011). The aim is more to ensure a realistic and fair assessment rather than to facilitate a comprehensive formative assessment that can trigger timely and appropriate teacher intervention during the learning process. The second stream focuses on developing systems exploiting Wiki data to inform continual assessment (e.g., Kubincová, Homola, & Janajev, 2012; Palomo-Duarte, Dodero, Medina-Bulo, Rodríguez-Posada, & Ruiz-Rube, 2014). These studies tend to be platform-specific and quantitative in nature, with a scope largely confined to the page edits. While useful, they fall short in providing comprehensive information about the quality of student learning.

It is important to be aware that student collaboration resides in the composite of activities enabled in Wiki. Studies found that page commenting may help promoting group collaboration and revealing group dynamics (Judd et al., 2010). Woo et al. (2013) further indicated that page comments were sometimes “revision-oriented” triggering page edits that led to enhancement of writing quality. It is therefore necessary to extend the focus from page contents and edits to page comments for a more all-round review of student activities in Wiki.
Student online discussions

Student comments in Wiki are similar to discussions on online forums to some extent, especially those forums designed for inquiry-based collaborative learning (Tirado, Aguaded, & Hernando, 2011). While the former is rarely studied, numerous studies have been conducted to investigate the dynamics of student online discussions. Many of them analyzed the content of student postings by categorizing them according to a range of dimensions. Pena-Shaff and Nicholls (2004) developed a framework of interactive learning to encode student postings in an online discussion board into “meaning construction” categories such as clarification, elaboration and interpretation. They also considered the interactivity among students and differentiated individual reflections from conversational interactions. Tirado et al. (2011) analyzed the activity records of an online discussion forum from the perspectives of “psycho-social relations,” “positive interdependence” and “construction of meaning.” Similarly, Xie and Ke (2011) conducted a content analysis of student online discussions in terms of “social interaction,” “knowledge construction” and “regulation of learning” (i.e., the coordination and management of the collaborative learning process). They observed that students tended to be involved in lower order cognitive activities when interacting with others but higher order cognitive activities when working on their own. In So’s study (2009), student online discussions during group projects were analyzed with regard to collaborative learning behaviors and social presence behaviors. They found that most discussions were about group work facilitation while relatively few activities were challenging and explaining/elaborating.

These studies considered both social and cognitive dimensions of student interactions which also reside in Wiki commenting. These dimensions serve well as the theoretical basis for this study. However, unlike online forums where the discussion itself is the expected artefact, Wiki commenting is to facilitate the co-construction of the Wiki page contents. Therefore, besides idea exchange and negotiation, it is essential for Wiki comments to be (1) able to help sustain group interaction and collaboration (Judd et al., 2010), (2) relevant and contributive to the overall discussion about the project in question, and (3) able to reflect the quality of thinking of the commenter (Woo et al., 2013). In addition, the commenting area of Wiki platforms are often designed much simpler than fully-fledged discussion forums (e.g., without subject lines). With these differences on purposes, functionalities and interfaces, whether and to what extent the existing frameworks developed for online forum discussions can be used to effectively analyze Wiki comments remain open questions. In this study, therefore, we draw from these studies to develop a multi-focal framework for page comment analyses adequate and adaptable at different project phases.

Methods

This study adopts an iterative approach using a combination of literature review, stakeholder interview, experimental coding, statistical analysis and association rule mining, a data mining approach to find out relationships between items or categories in a dataset. An initial framework was first developed from the literature (FW V0) and then modified based on interviews with in-service teachers (FW V1). A sample of students’ page comments on group project Wikis were collected and coded with the refined framework (FW V1). The distribution of the coded data informed yet another round of framework revision (FW V2). Association rule mining was then applied to the data coded with FW V2 to find out relationships among the categories. The results informed the final round of framework revision (FW V3).

Initial framework development

According to the three requirements mentioned in the previous section, related frameworks and/or taxonomies on the following three aspects in the literature of student online asynchronous discussions were adopted: Social Interaction (Bales, 1950; Tirado et al., 2011), Meaning Construction (Pena-Shaff & Nicholls, 2004), and Thinking Development (Krathwohl, 2002). A composite framework (FW V0) consisting of three modules (i.e., Social Interaction, Meaning Construction and Thinking Development) was consequently developed and presented in the Appendix.

Interviews with in-service teachers

To examine the practicality of this initial framework and to further refine it, the opinions of seven in-service secondary school teachers were sought via five semi-structured interviews. Convenience sampling was adopted to invite the interviewees. Among them, four were invited to inform the refinement (i.e., “exploration group”)
and three to comment on the refined scheme to ensure its feasibility and practicality (i.e., “evaluation group”). Both groups were asked (1) what they would need to know about student comments in Wikis; (2) how they thought of the framework in relation to their practical needs; and (3) how the framework can be modified. The evaluation group was also asked to comment on the refined framework developed during the exploration stage. Because of the difference in purpose, interviews were conducted in pairs with the exploration group to allow collective brainstorming, discussion and a certain extent of consensus building. Interviews with the evaluation group were conducted individually to ensure independent judgment. Table 1 summarizes the demographic characteristics of the interviewed teachers. The refined framework after the interviews is referred to as FW V1 and is presented in the Appendix.

Table 1. Demographic information of the interviewees

<table>
<thead>
<tr>
<th>Gender</th>
<th>Exploration group</th>
<th>Evaluation group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Female</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Teaching experience</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-3 years</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8-10 years</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Over 10 years</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Subject fields</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Humanities (Chinese, English, Design)</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Social Science (History)</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Science (Chemistry, Computer, Geography)</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Note. (1) One interviewee teaches more than one subject.

Experimental coding

To test the applicability of FW V1, we applied it to a sample of student comments collected from the Wikis developed by 48 student groups from a junior secondary school in Hong Kong. Of the 48 groups, 30 were from Form 1 (equivalent to Grade 6) and 18 were from Form 2 (Grade 7). The total number of students involved was 238. Their age ranged from 12 to 14 during the study period. Each group consisted of about five students who were required to collaborate on an inquiry-based project for the Liberal Studies course on a five-month period. A Google Site was created exclusively for each group to facilitate collaboration. The groups were required to write their project reports on their Google Sites using Wiki pages to differentiate sections (e.g., introduction, methodology). During the project period, the students could post comments on each page and use this feature to communicate and discuss with one another. For this study, the comments attached to each page, which are predominantly written in English, were collected. Among the 48 Wikis, 7 contained no comments (6 from Form 1 and 1 from Form 2) and 1 had access control in place. They were thus removed from the sample.

The unit of analysis varied in existing studies involving content analysis of student online discussions, including unit of meaning (Bales, 1950), sentence (Pena-Shaff & Nicholls, 2004) and postings (Weltzer-Ward, 2011). To sustain the comparability across modules, sentence was used in this study as the unit of analysis. In case of grammatical irregularities, an operational definition of sentence, a set of words resembling a simple and coherent utterance, was adopted. Table 2 provides a descriptive summary of the sample that consists of 1,482 units.

Table 2. A descriptive summary about the dataset

<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th>Form 1</th>
<th>Form 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of units</td>
<td>1,482</td>
<td>1,056</td>
<td>426</td>
</tr>
<tr>
<td>Number of groups</td>
<td>48</td>
<td>30</td>
<td>18</td>
</tr>
<tr>
<td>Number of groups to be studied</td>
<td>40</td>
<td>23</td>
<td>17</td>
</tr>
</tbody>
</table>

All units identified were coded manually according to FW V1 and the results were used to inform further refinement of the framework to FW V2. To ensure coding quality, comments from one fourth of student groups randomly selected in both Forms were double-coded by a second independent coder. Cohen’s Kappa was calculated to measure the level of interrater reliability. The distribution of the coded data across different categories were then analyzed and compared to findings in related literature.
Relationship between comment categories

Association rule mining is a data mining technique used for identifying associations among frequently appeared patterns in a dataset (Han, 2012). It has been used in the education domain to find out relationships between variables, particularly in datasets with many variables (e.g., student emotion status, learning performances) (Baker, 2010). Unlike correlation analysis that is bivariate, association rule mining can discover relationships among multiple variables at the same time. Specifically, association rule mining aims to find “if-then” rules of the variables, in the form of “antecedent $\rightarrow$ consequence,” where antecedent and consequence are conditions that some variable(s) has certain value(s). This study applied association rule mining to explore the associations among categories across the three modules. For example, a possible rule in this study might be “a comment is in SI-0 $\rightarrow$ the comment is in MC-0.” That is, if a comment is categorized into the SI-0 category, chances are it is also in the MC-0 category.

To identify interesting and significant rules, the FP-Growth algorithm (Han, 2012) was used with a minimum Support value of 0.50 and minimum Lift value of 2.0. Rules satisfying the criteria are defined as “interesting” ones (Han, 2012). In addition, the “Cosine” measures of interestingness proposed by Merceron and Yacef (2008) for association rules in educational data were adopted. The “Cosine” interestingness threshold is set on the level larger than 0.65. These threshold values were set with tradeoffs between frequency of occurrences and number of resultant rules.

Results

Refined framework FW V1

In view of FW V0, the interviewees in the exploration group agreed that the three modules in the framework were necessary. However, they were concerned of the categorization complexity, which may turn out to be impractical as it may take much time for teachers to understand the categories and interpret the results. Some categories were regarded rare among their students. An example was the category “reflection,” which they thought students would seldom do upon commenting. The interviewees also agreed that it was necessary to combine categories similar in nature to make the framework more practically feasible. It was noted that more detailed categorization would be feasible or needed either when more resources (e.g., time and manpower) are made available to teachers or when student comments reflect a strong inclination toward a particular category.

To refine the scheme, possible code combinations were proposed to solicit discussion within the two exploration groups who also suggested combinations they found fit for their needs. Consensus was reached after each pair interview and the opinions from the interviews were consolidated to inform framework refinement. The refinement decisions made at this stage are summarized as below.

In Social Interaction, three sets of codes were combined due to their shared natures, namely, (1) SI-1 to SI-3 (Give Suggestions/Information/Opinions) being combined into SI-A (Giving Acts) as they represents different types of giving acts; (2) SI-4 to SI-7 (Ask for Suggestions/Information/Opinions/Help) being combined into SI-B (Make Requests) as they represents different types of requests; and (3) SI-11 (Encouragement) and SI-12 (Others) being combined into SI-E (Others) as they both concern students’ ability in socializing. There were different opinions on whether SI-8 (Agree) and SI-9 (Disagree) should remain as individual categories. When commenting on the Meaning Construction module, one interviewee noted that student conflicts (i.e., MC-5) may need teachers’ special intervention. In line with this opinion, these two categories were kept separate while SI-9 (Disagree) was combined with SI-10 (Show Antagonism) into SI-D (Disagree) as they are both acts of negating others.

In Meaning Construction, six codes, namely, MC-2 (Reply), MC-3 (Clarification), MC-4 (Interpretation), MC-6 (Assertion), MC-8 (Judgment), and MC-10 (Support), which involve acts of responding (to previous questions, statements or perspectives) or providing feedback were combined into MC-X (Responses). MC-7 (Consensus Building) and MC-9 (Reflection) were combined into MC-D (Result) as they indicate some kind of results (of a discussion and/or a learning process).

In Thinking Development, the interviewees were asked whether they preferred two categories (i.e., either there is cognitive development occurred or not) or three categories (i.e., no, low or high level of cognitive development). They tended to find dualism arbitrary and were inclined to have three categories. So TD-1 to TD-3 (Remember,
Understand, and Apply) were combined into TD-L representing low level of cognitive development, and TD-4 to TD-6 (Analyze, Evaluate, and Create) into TD-H representing high level of cognitive development.

The feasibility and practicality of the consequent refined framework (FW V1) were confirmed in the evaluation group interviews. The interviewees generally expressed that they would be more willing to use the refined framework and knowing the overall picture would be sufficient in view of their heavy workload. One interviewee suggested the Thinking Development module be divided into four categories (no, low, medium or high cognitive development). This suggestion was kept for further consideration in the next round of framework refinement.

**Trial coding and FW V2**

FW V1 was tried out with the page comments collected from the student Wikis, to verify the framework and further refine it to FW V2. To ensure coding reliability, 189 randomly selected units were coded with two independent coders. Cohen’s kappa reached 0.821, 0.839 and 0.848 for Social Interaction, Meaning Construction and Thinking Development respectively, indicating very good interrater reliability (Altman, 1991). The refinement decisions made at this stage are summarized below.

In Social Interaction, the statistics of the coded data indicated SI-A (Giving Acts) accounted for about 70% of the comment units. To avoid such strong inclination, SI-A (Giving Acts) was segregated back to SI-1 (Give Suggestions), SI-2 (Give Opinions) and SI-3 (Give Information) in FW V2.

In Meaning Construction, the statistics of the coded data indicated MC-X (Responses) accounted for half of the comment units. Thus, the categories included in MC-X were rearranged in FW V2: MC-2 (Reply) was divided into two according to the complexity of the replies. The type of responses that are simple and straightforward in nature was combined with MC-10 (Support) to form the new category MC-B (Simple Replies). Those that are not simple was combined with MC-3 (Clarification), MC-4 (Interpretation), MC-6 (Assertion) and MC-8 (Judgment) to form another new category, MC-F (Argument) to capture comments that are argumentative in nature.

In Thinking Development, the question was whether to have three or four categories. The statistics of the coded data indicated the comment units reflecting the presence of cognitive development are of the small minority (around 30%). The small sample size makes it hard to judge whether finer categorization would enhance the expressiveness of the results. Therefore, no changes were made and three categories were kept in this module.

The framework developed up to this stage (FW V2) is presented in the Appendix. In comparison with those commonly adopted in online discussion research, this framework emphasizes teachers’ practical needs, and the three modules are particularly helpful in supporting separate analyses of student comments from different perspectives of concern.

The distribution of the coded data across the categories in FW V2 is shown in Table 3. It is noteworthy that both Forms have similar distributions. For both Forms, student comments were dominated by two to three frequent categories in each module. In Social Interaction, over 70% of the units fall under the three categories of giving acts (i.e., SI-1: Give suggestions, SI-2: Give opinions, and SI-3: Give information). In Meaning Construction, MC-0 (N/A) and MC-F (Argument) were most frequent in both Forms, with each accounting for about 30% of all comments. In Thinking Development, the dominance of TD-0 (N/A) is obvious, followed by TD-H (High Cognitive Development).

The interaction/discussion patterns as reflected in the results are comparable to the findings of So (2009), where frequent online discussion activities among students were noted to include organizing work, reporting progress, sharing task, feedback seeking/giving, and group cohesion. These activities are related to the popular categories in the coded sample (Table 3) including SI-3 (Giving Information), MC-0 (N/A) and TD-0 (N/A). Different from Xie and Ken’s findings (2011) that there tended to be more lower-level than higher-level knowledge construction behaviors in online discussion, there were more TD-H (High Cognitive Development) than TD-L (Low Cognitive Development) observed in the present study. This is possibly due to the different settings of the two studies. In Xie and Ke (2011), students were required to share information and participate in discussions as part of the course assessment whereas in the present study commenting was optional and was not the only channel for information sharing and interaction. This might have affected students’ motivation in posting and the consequent interaction behaviors and discussion patterns.
Table 3. Distribution of coded categories within each module in FW V2

<table>
<thead>
<tr>
<th>Module</th>
<th>Category</th>
<th>Form 1</th>
<th>Form 2</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social interaction</td>
<td>SI-0: N/A</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>SI-1: Give Suggestions</td>
<td>25%</td>
<td>23%</td>
<td>25%</td>
</tr>
<tr>
<td></td>
<td>SI-2: Give Opinions</td>
<td>26%</td>
<td>28%</td>
<td>27%</td>
</tr>
<tr>
<td></td>
<td>SI-3: Give Information</td>
<td>21%</td>
<td>21%</td>
<td>21%</td>
</tr>
<tr>
<td></td>
<td>SI-B: Ask</td>
<td>12%</td>
<td>9%</td>
<td>11%</td>
</tr>
<tr>
<td></td>
<td>SI-C: Agree</td>
<td>5%</td>
<td>5%</td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>SI-D: Disagree</td>
<td>4%</td>
<td>3%</td>
<td>4%</td>
</tr>
<tr>
<td></td>
<td>SI-E: Others</td>
<td>7%</td>
<td>10%</td>
<td>8%</td>
</tr>
<tr>
<td>Meaning construction</td>
<td>MC-0: N/A</td>
<td>30%</td>
<td>20%</td>
<td>27%</td>
</tr>
<tr>
<td></td>
<td>MC-A: Questions</td>
<td>13%</td>
<td>12%</td>
<td>13%</td>
</tr>
<tr>
<td></td>
<td>MC-B: Simple Replies</td>
<td>10%</td>
<td>13%</td>
<td>11%</td>
</tr>
<tr>
<td></td>
<td>MC-C: Conflict</td>
<td>4%</td>
<td>2%</td>
<td>3%</td>
</tr>
<tr>
<td></td>
<td>MC-D: Results</td>
<td>5%</td>
<td>0%</td>
<td>4%</td>
</tr>
<tr>
<td></td>
<td>MC-E: Others</td>
<td>4%</td>
<td>6%</td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>MC-F: Argument</td>
<td>33%</td>
<td>47%</td>
<td>37%</td>
</tr>
<tr>
<td>Thinking development</td>
<td>TD-0: N/A</td>
<td>68%</td>
<td>66%</td>
<td>67%</td>
</tr>
<tr>
<td></td>
<td>TD-L: Low Cognitive Development</td>
<td>7%</td>
<td>6%</td>
<td>7%</td>
</tr>
<tr>
<td></td>
<td>TD-H: High Cognitive Development</td>
<td>25%</td>
<td>27%</td>
<td>26%</td>
</tr>
</tbody>
</table>

Note. *Categories SI-1, SI-2, SI-3, MC-B and MC-F were recoded by two coders. The interrater reliability are: SI: $\kappa = 0.872$; MC: $\kappa = 0.851$; TD: $\kappa = 0.848$.

Relationships of codes

Table 4 shows strong association rules mined from the coded dataset. All of them have Lift values larger than 2.00. Those with cosine values larger than 0.65 are recognized as very strong rules according to Merceron and Yacef (2008).

<table>
<thead>
<tr>
<th>Antecedent</th>
<th>Consequence</th>
<th>Support</th>
<th>Confidence</th>
<th>Lift</th>
<th>Cosine</th>
</tr>
</thead>
<tbody>
<tr>
<td>TD-0, SI-1</td>
<td>MC-0</td>
<td>0.13</td>
<td>0.61</td>
<td>2.21</td>
<td>0.54</td>
</tr>
<tr>
<td>SI-3</td>
<td>TD-0, MC-0</td>
<td>0.13</td>
<td>0.61</td>
<td>2.24</td>
<td>0.54</td>
</tr>
<tr>
<td>SI-2</td>
<td>MC-F, TD-H</td>
<td>0.15</td>
<td>0.56</td>
<td>3.01</td>
<td>0.67</td>
</tr>
<tr>
<td>TD-H</td>
<td>MC-F, SI-2</td>
<td>0.15</td>
<td>0.58</td>
<td>2.65</td>
<td>0.63</td>
</tr>
<tr>
<td>MC-F</td>
<td>SI-2</td>
<td>0.22</td>
<td>0.6</td>
<td>2.25</td>
<td>0.71</td>
</tr>
<tr>
<td>MC-F, SI-2</td>
<td>TD-H</td>
<td>0.15</td>
<td>0.68</td>
<td>2.65</td>
<td>0.63</td>
</tr>
</tbody>
</table>

The first two rules in Table 4 involve the associations among SI-1 (Give Suggestions), SI-3 (Give Information), MC-0 (N/A) and TD-0 (N/A). In particular, the rule “SI-3 -> MC-0, TD-0” suggests that comments in SI-3 (Give Information) are likely to be in MC-0 (N/A) and TD-0 (N/A) as well. The next four rules indicate that the interrelationships between SI-2 (Give Opinions), MC-F (Argument) and TD-H (High Cognitive Development) are strong and reciprocal, suggesting that comments in each of these categories are likely to be in the other two categories as well (except for MC-F -> SI-2, TD-H).

In line with these views that SI-1 (Give Suggestions) and SI-3 (Give Information) had similar implications, to enhance the filtering capacity of the scheme, SI-1 and SI-3 were combined, leading to FW V3 as shown in the Appendix.

Discussions

Results of this study have implications in teaching practice. The fact that most comments were in MC-0 (N/A) and TD-0 (N/A) categories suggests that the comment area in Wiki platform can be enhanced. Compared to fully-fledged discussion forums where deep discussions are more often, the Wiki comment areas usually have fewer functionalities and limited affordance. For example, the comment area in Google Sites where the sample data were collected did not support multiple discussion threads, subject line, or indentation of replies. Thus it was difficult to tell the discussion structure (i.e., who responded to whom). Similar issues exist on other Wiki
platforms often used in schools including PBworks, Wikispaces and the Wiki activity in Moodle, one of the most popular Learning Management Systems. In contrast, MediaWiki, the platform used by Wikipedia, has a more flexible discussion area that supports multiple threads, heading for each thread, indentions within each thread and the option of collapsing/expanding arbitrary levels of replies. As a result, many in-depth discussions on complex topics such as computer programming and knowledge engineering are supported in MediaWiki (Di Francescomarino, Ghidini & Rospocher, 2012). However, it is noteworthy that more structure built in the Wiki comment area may not necessarily bring in more benefits. After all, the purpose of Wiki commenting is to facilitate collaborative writing, while the discussion itself is not the targeted outcome. Therefore, a semi-formal approach to discussion could be more proper. Complex structures such as reference links and “build-on” relationships in the Knowledge Forum (Yang, van Aalst, Chan, & Tian, 2016) might possibly impose unnecessary cognitive load to the students, even though they can greatly facilitate collaborative knowledge co-construction and creation in the knowledge building context.

The dominance of MC-0 and TD-0 also suggests that interventions are desirable during the project period. Analytic tools can be developed using the developed framework to help teachers check student comments regularly. Early intervention can then be implemented once insufficient higher-level comments are detected. Similar tools can also be used by students as a means of obtaining timely feedback that can facilitate self-regulated learning (Carless, Salter, Yang & Lam, 2011).

One of the important goals of education is to help students acquire higher-order cognitive skills as embodied by TD-H (High Cognitive Development) in the framework. As such, the association rules found (Table 4) suggest that giving information (SI-1) or suggestions (SI-3) only may not need to be particularly encouraged during collaborative projects whereas it is worthwhile to encourage students to post opinionated comments (SI-2), to present arguments (MC-F), or both in order to increase the chance of using higher-order cognitive skills (TD-H).

The relationships among categories also support the possibility of using the outcomes from one module to predict the general outcomes of the other two. For instance, an observed inclination toward SI-1 (Give Suggestions) and SI-3 (Give Information) can be taken as indicators of the possible low frequencies of MC-F (Argument) and TD-H (High Cognitive Development). In this way, schools may selectively implement one module instead of the full framework, depending on the workload of their teachers and resources available. This reflects the flexibility of the framework.

Last but not least, the similar distribution of categories across the two Forms (Table 3) can be exploited to build automated categorization tools (Rosé et al., 2008). Although Form 1 groups had more MC-0 (N/A) whereas Form 2 groups had more MC-F (Argument), results of Mann-Whitney U Tests (as the sample did not follow normal distribution) showed no significant difference between the two Forms throughout all categories. Comments collected from one Form can be manually annotated and used to train the categorization model. The model then can be applied to comments of the other Form to automatically generate the categories.

**Conclusion and future work**

Student comments in Wikis are important evidence revealing students’ ongoing collaborative writing processes. The goal of this study is to develop a framework for analyzing students’ page comments posted in Wikis of group writing to inform affirmative learning assessment and feedback. A three-module framework was developed from current literature and enhanced with interviews of in-service teachers. The refined scheme was tested with real-life data collected from group Wikis of students in a secondary school in Hong Kong. The coded sample was analyzed statistically in terms of distributions. Relationships across categories in different modules were also detected with association rule mining. The final product is, therefore, verified to be practically useful in providing readily interpretable information about the quality of student interaction, meaning construction and thinking development as reflected in their Wiki page comments.

This study has several limitations. First, it is contextually confined to Hong Kong where student behaviors may be distinctive as a consequence of its socio-cultural uniqueness. Second, this study was informed by data from junior secondary school students from one local school in Hong Kong. The results may be different among other student populations (e.g., senior secondary school students, students in international schools). Third, while Hong Kong is a multilingual city, a majority of its population speaks Cantonese. The data of this study comes from native speakers of Cantonese. The comment pattern might be different if they used Chinese rather than English predominantly in their comments. Therefore, more empirical studies are necessary to test the applicability of the findings in this study to other regions, other languages and other student populations.
Undoubtedly, this study focuses only on one facet of student activities in Wiki. As future work, this study will be further developed in conjunction with existing studies on Wiki page edits for a more comprehensive model of assessing Wiki activities. In addition, the evolution of student commenting behaviors during the project period may disclose meaningful patterns of the learning process and is worth further investigation.

Acknowledgements

This study was supported by an Early Career Scheme grant from the Research Grants Council of the Hong Kong Special Administrative Region, China. (Project No. HKU 27401114) and the Seed Funding Programme for Applied Research in the University of Hong Kong.

References


### Appendix: Coding Scheme Evolution*

<table>
<thead>
<tr>
<th>Source</th>
<th>V0 category</th>
<th>V1 category</th>
<th>V2 category</th>
<th>V3 category</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Social Interaction (SI)</strong></td>
<td><strong>Objective:</strong> To measure students’ efforts devoted to maintaining and sustaining the interaction and collaboration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bales (1950)</td>
<td>SI-0: N/A No social interaction reflected</td>
<td>SI-0</td>
<td>SI-0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SI-1: Give suggestions</td>
<td>SI-A: Giving acts</td>
<td>SI-1</td>
<td>SI-A</td>
</tr>
<tr>
<td></td>
<td>Propose directions, possible ways of actions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SI-3: Give information (1)</td>
<td></td>
<td>SI-3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inform, repeat, clarify, confirm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SI-2: Give opinions</td>
<td></td>
<td>SI-2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Express thoughts, feelings, wishes; evaluate; analyze</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SI-4: Ask for information (1)</td>
<td>SI-B: Make requests</td>
<td>SI-B</td>
<td>SI-B</td>
</tr>
<tr>
<td></td>
<td>Request for information, repetition, clarification, confirmation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SI-5: Ask for opinions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Request for thoughts, expressions of feelings and wishes, evaluation, analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SI-6: Ask for suggestions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Request for directions, possible ways of action</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SI-7: Ask for help (2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Request for support and assistance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SI-8: Agree</td>
<td>SI-C: Agree</td>
<td>SI-C</td>
<td>SI-C</td>
</tr>
<tr>
<td></td>
<td>Show passive acceptance, understanding, compliance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SI-9: Disagree</td>
<td>SI-D: Disagree</td>
<td>SI-D</td>
<td>SI-D</td>
</tr>
<tr>
<td></td>
<td>Show passive rejection, withhold help</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SI-10: Show antagonism</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Deflate other’s status, defend or assert self</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cheer up other, express support, aiming to promote group cohesion</td>
<td>Building comradeship, socialize</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SI-12: Others (3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Social statements such as greetings and apologies that facilitates communication</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

* denotes evolution from previous coding schemes.
### Meaning Construction (MC)

**Objective:** To measure the contribution and relevancy of the comments to the overall discussion about the project in question

<table>
<thead>
<tr>
<th>Pena-Shaff &amp; Nicholls (2004)</th>
<th>MC-0: N/A Comments with no observed contribution / relevancy to the discussion</th>
<th>MC-0</th>
<th>MC-0</th>
<th>MC-0</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC-1: Question</td>
<td>Gather unknown information, inquire, start a discussion, reflect on problems raised</td>
<td>MC-A: Question</td>
<td>MC-A</td>
<td>MC-A</td>
</tr>
<tr>
<td>MC-2: Reply</td>
<td>Respond to or elaborate on other’s questions or statements</td>
<td>MC-X: Response Provide feedback or address previous questions, statements or perspectives</td>
<td>MC-B: Simple Replies Make simple responses to other’s questions or statements, establish rapport, share feelings, agree with other’s ideas either directly or indirectly</td>
<td>MC-B</td>
</tr>
<tr>
<td>MC-3: Clarification</td>
<td>Identify and elaborate on ideas and thoughts</td>
<td></td>
<td>MC-F: Argument Make statements or responses that are argumentative in nature or contribute to the argument for or against a thought or idea</td>
<td>MC-F</td>
</tr>
<tr>
<td>MC-4: Interpretation</td>
<td>Analyze, make predictions, build hypotheses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC-5: Conflict</td>
<td>Debate other’s viewpoints, show disagreement, show friction with others</td>
<td>MC-C: Conflict</td>
<td>MC-C</td>
<td>MC-C</td>
</tr>
<tr>
<td>MC-6: Assertion</td>
<td>Provide explanations and arguments to maintain and defend ideas questioned by other</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC-7: Consensus building</td>
<td>Try to attain a common understanding of the issues in debate</td>
<td>MC-D: Result Conclude a discussion, reflect upon the learning process of the project</td>
<td>MC-D</td>
<td>MC-D</td>
</tr>
<tr>
<td>MC-9: Reflection</td>
<td>Acknowledge learning something new, judge importance of discussion topic in relation to learning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Revised Bloom’s Taxonomy in the cognitive domain (Krathwohl, 2002)</td>
<td>Thinking Development (TD)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MC-E: Others</strong></td>
<td><strong>Objective:</strong> To measure the quality of student learning as reflected in their cognitive development</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC-E</td>
<td><strong>MC-E:</strong> Others</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Social statements, messages with functions unable to categorize</strong></td>
<td><strong>MC-E:</strong> Others</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Thinking Development (TD)

**TD-0:** N/A  
No cognitive development observed

**TD-1:** Remember  
Retrieving relevant knowledge from long-term memory

**TD-2:** Understand  
Determining the meaning of instructional messages, including oral, written, and graphic communication

**TD-3:** Apply  
Carrying out or use a procedure in a given situation

**TD-4:** Analyze  
Breaking material into its constituent parts and detecting how the parts relate to one another and to an overall structure or purpose

**TD-5:** Evaluate  
Making judgments based on criteria and standards

**TD-6:** Create  
Putting elements together to form a novel, coherent whole or make an original product

**TD-L:** Low cognitive development  
- Retrieving relevant knowledge from long-term memory  
- Determining the meaning of instructional messages, including oral, written, and graphic communication  
- Carrying out or use a procedure in a given situation

**TD-H:** High cognitive development  
- Breaking material into its constituent parts and detecting how the parts relate to one another and to an overall structure or purpose  
- Making judgments based on criteria and standards  
- Putting elements together to form a novel, coherent whole or make an original product

**Notes:**

1. The categories “Give Orientation” and “Ask for Orientation” in Bales’ categorization (1950) were rephrased into “Give Information” and “Ask for Information” to highlight the act of information exchange seemingly central to a constructivist learning process.
2. The scope of the category ‘Show Tension’ in Bales’ categorization (1950) were narrowed down to ‘Ask for Help’ to highlight the act of seeking help that is commonly seen in a learning process.
3. These two categories are adapted from Tirado, Aguaded, & Hernando (2011) to capture acts of team building and socialization that are deemed essential to facilitate communication and collaboration.

* The category definitions are adapted from the designated sources with slight modifications from the present authors. Examples of categories are presented in this linked document:  
http://ccmir.cite.hku.hk/data/FW_examples.pdf
Strengthening Social Networks in Online Discussion Forums to Facilitate Help Seeking for Solving Problems

Po-Yao Chao1, K. Robert Lai*, Chen-Chung Liu2 and Hung-Ming Lin3

1Yuan Ze University, Taiwan // 2National Central University, Taiwan // 3Minghsin University of Science and Technology, Taiwan // poyaochao@saturn.yzu.edu.tw // krlai@cs.yzu.edu.tw // ccliu@cl.ncu.edu.tw // hmlin@must.edu.tw

*Corresponding author

(Submitted August 18, 2016; Revised April 18, 2017; Accepted July 1, 2017)

ABSTRACT
Help seeking is regarded as an important learning strategy that reflects students’ metacognitive and domain-specific skills and knowledge. This study developed a proactive online discussion forum that strengthens social networks in an online discussion forum to facilitate help seeking for problem solving. This forum allowed students to proactively invite friends or potential experts to participate in problem solving. A quasi-experimental approach was conducted with a sample of 37 students. The students’ behaviors of problem solving, ego network patterns of problem solving, and perceptions of help seeking were addressed as dependent variables to assess the effect of the proactive online discussion forum. The analysis results revealed that the students in the experimental group exhibited more frequent participation in problem solving, and showed more positive attitudes towards help seeking. The results also indicated that the experimental group demonstrated more in-depth participation patterns of ego networks. The findings provide some evidence for the effectiveness of the proactive online discussion forum in terms of promoting students’ participation in problem solving and their attitudes towards help seeking.

Keywords
Online discussion forum, Ego network, Help seeking, Online social network

Introduction
Help seeking has been advocated as an important learning strategy that reflects students’ metacognitive and domain-specific skills and knowledge (Aleven, Stahl, Schworm, Fischer, & Wallace, 2003; Newman, 2002). It involves identifying a problem, expressing the need for help, and receiving assistance from others (Newman, 2002; Schunk, 2004). Among the forms of help-seeking, students asking for help in order to learn independently and exhibit self-determination to solve their problems could benefit self-regulation and learning outcomes (Cheng & Tsai, 2011; Melrose, 2006). However, during typical academic terms, the majority of college students reported that they faced challenges or difficulties when seeking help with academic problems (Karabenick & Knapp, 1988). In such cases, some students may exert little effort to solve their academic problems, give up prematurely, or persist unsuccessfully on their own (Newman, 2002). For help seekers, the costs and threat of seeking needed help from others becomes a major determinant of their help-seeking intentions, their help-seeking purposes, and their preferred sources of assistance (Karabenick, 2003). The students’ help seeking difficulties may emphasize the cost and threat of help seeking, which may in turn initiate the avoidance of help seeking or may lead to executive help seeking behaviors in which students’ intention is to have someone else solve a problem on their behalf (Karabenick & Knapp, 1991). Given that students are encouraged to be more autonomous and responsible for their learning (Schunk, 2004), their help-seeking difficulties may be a barrier to acquiring needed resources during the process of solving academic problems.

Research shows that the learning environments that assist students to identify their problems and seek help from capable peers may enhance students’ motivation to study, or may promote their performance in studying subjects (Melrose, 2006). Among the learning environments, online discussion forums such as WebCT or phpBB, which have been widely incorporated in courses to help students solve problems, could enhance the process of acquiring, sharing and exchanging knowledge among students, and hence may improve the process of help seeking and learning outcomes (Cheng, Pari, Collimore, & Joordens, 2011). Although online discussion forums have demonstrated potential for supporting students to help one another in the resolution of academic problems, the opportunity and quality of learning through online discussion largely depends on student participation and meaningful interaction among the members of the forum (Cheng et al., 2011; Guzdial & Turn, 2000; Hew & Cheung, 2008). Moreover, research has shown that only a few students would like to share their questions with others or to give comments on questions in online discussion forums (Wan & Johnson, 1994). The lack of knowledge sharing in online discussion forums may be due to the dilemmas of knowledge sharing (Cabrera & Cabrera, 2002). One of the dilemmas is that individuals attempt to maximize their self-interests and benefits in a...
community, which makes them inclined not to contribute and can consequently lead to collective damage (Cabrera & Cabrera, 2002). In such dilemmatic cases, the interdependent relationship that requires joint cognitive and emotional engagement by help-seeker and help-giver(s) hardly emerges naturally. Networks that maintain connections among relevant people (e.g., friends or family members) may provide a basic structure or potential helpers to promote such seeker-help giver relationship. In addition, the help seekers’ decision to ask questions and the help givers’ willingness to respond to the questions play significant roles in successful help seeking processes (Gall, 1985). Difficulties such as help seekers’ inability to ask for help or help givers’ perceived irrelevance to the seekers’ help may thus impede students’ participation in online discussion (Balaji & Chakrabarti, 2010; Guzdial & Turns, 2000), and therefore hinder students’ engagement in help seeking processes such as giving help or receiving assistance from others in online discussion forums.

The use of social networking sites (e.g., Facebook or MySpace) has become increasingly ubiquitous among students. These sites assist users in connecting with relevant people (e.g., friends, family members or classmates) and in sharing information with these people (Hew, 2011; Lockyer & Patterson, 2008). In these social networking sites, the connections between a person and relevant others (e.g., friends or family members) are maintained by online social networks, which can serve as a major resource for a person to seek assistance (Gourash, 1978; Park, Lee, & Kim, 2012). The online social networks sustaining meaningful relationships among connected people have the potential to facilitate students’ help seeking and giving for solving problems in online discussion forums. Given the potential of the online social networks and the difficulties that hinder students’ engagement in help seeking processes, the purposes of this study were twofold. First, a proactive online discussion forum that strengthens online social networks in an online discussion forum to facilitate help seeking for problem solving was developed. Instead of passively waiting for responses from others, the proactive online discussion forum allowed individual learners to proactively send messages to friends or potential experts to invite advice, and to maintain a social network for solving problems. The second purpose was to explore the effect of the proactive online discussion forum on students’ participation in problem solving and students’ perceptions of help seeking with the proactive online discussion forum.

Related work

Help seeking with online discussion forum systems

An online discussion forum system is a computer-based application that provides an online learning environment to support discussion and debate relevant to the course materials among students and teachers (Cheng et al., 2011; Guzdial & Turns, 2000; Hew & Cheung, 2008). Several studies have adopted online discussion forum systems to support students in seeking help online for academic problem solving (e.g., Balaji & Chakrabarti, 2010; Bull, Greer, McCalla, & Kettel, 2001). To solve their problems with online discussion forums, help seekers should formulate questions based on their problems and expect that potential helpers such as capable peers or teachers would give useful advice regarding their questions. In such situations, the help seekers’ (questioners’) and help givers’ (peers’) participation is a requirement of the problem solving collaboration (Bull et al., 2001; Cheng et al., 2011; Guzdial & Turns, 2000; Hew & Cheung, 2008). However, motivating students to participate in and contribute to online discussion for help seeking and giving help is challenging (Balaji & Chakrabarti, 2010). According to Cabrera and Cabrera’s (2002) knowledge sharing dilemmas, individuals may tend not to share their knowledge in a virtual community because the process of knowledge sharing may involve significant costs (e.g., making that knowledge available to others) and, more important, because individuals’ refusal to share knowledge with others may appear to be the most advantageous strategy in securing their benefit within a knowledge sharing context (Liu, Lin, Chang, & Chao, 2014). Considering such dilemmatic situation, resolving the dilemmas of knowledge sharing becomes a major issue of assisting help seeking in an online discussion forum. Social capital theory provides some viewpoints that may contribute to the resolution of knowledge sharing dilemmas: establishing or utilizing the relationship of trust among community members to assist help seeking and giving. The social capital theory stresses that an individual can draw on resources from other members of the network to which the individual connects (Ellison, Steinfield, & Lampe, 2007). In this sense, individual’s help-seeking in an online discussion forum can be improved if the members are willing to share knowledge or to interact with others. Inter-personal factors can influence the individuals’ willingness to share resources in virtual communities. For example, the experience of insecurity among members in an online learning environment may hinder members from sharing knowledge or participating in learning activities (Liu et al., 2014). On the other hand, mutual trust among members (e.g., friends) improves interactions and is critical to the success of knowledge sharing (Chow & Chan, 2008). Since it is important and useful for community members to seek help from more expert learners (Gall, 1985), recommending members with expertise to help seekers may also enhance the share of knowledge in virtual communities (Zhang, Ackerman, & Adamic, 2007).
Therefore, considering supports that invites trustworthy or competent members such as friends or expert learners to engage in help-seeking processes may improve members’ willingness to participate in discussion for problem solving in an online discussion forum. In addition, the difficulties or costs of seeking help in a learning environment may impact students’ attitudes toward, decision to, and purposes of seeking help (Karabenick, 2003). Hence, students’ perceptions of help seeking including instrumental or executive help-seeking and the threat or avoidance of help-seeking were employed in this study to explore the effects of an online discussion forum.

**Participation in academic problem solving using online discussion forums**

According to Wenger (1999), participation refers to “a process of taking part and also to the relations with others that reflect this process. It suggests both action and connections” (p. 55). In this regard, both the behavior of students’ problem solving and their connections with others could represent students’ participation in help seeking using an online discussion forum. Students’ problem solving in such an online discussion forum involves questioning, commenting on questions, and viewing comments. These activities are fundamental and crucial to the usefulness and effectiveness of the forum for students to solve their problems (Cheng et al., 2011; Hew & Cheung, 2008). Among the activities, questioning is considered a fundamental cognitive process that may be initiated by questioners when they suffer from contradictions, anomalous information, uncertainty, or obvious gaps in their knowledge (Otero & Graesser, 2001). It also serves as a strategy that an active help seeker would employ to elicit help (Nelson-Le Gall, 1981). With regard to giving comments to questions, the activity requires commenters to synthesize their idea for the comments and to phrase their idea, which could benefit learners’ construction of knowledge (Cheng et al., 2011; Guzdial & Turns, 2000). The viewing comments activity involves reading posts from others or from the readers themselves. It helps readers to keep up with the discussion or to monitor the status of interesting questions (Guzdial & Turns, 2000; Hew & Cheung, 2008).

Given the importance of problem solving activities, it is suggested that the core design of any online discussion forum should consider support for cultivating problem solving activities to facilitate students’ participation in help seeking.

**Social network analysis**

Social networks are believed to enable individuals to develop norms of trust and reciprocity, which are necessary for successful engagement in collective activities (Valenzuela, Park, & Kee, 2009). They could represent social connections between people and reflect the qualitative aspect of participation in collective activities. Social network analysis, which is designed to express patterns of relationships among members of social systems, could be adopted to assess students’ participation patterns in the online discussions (Rabany, Takaffoli, & Zaiane, 2012). Participants’ log files including information about the activity of the participants in the discussion forums are typically used to extract the social network underlying the discussion threads. Among the different types of social network, ego networks help understand variation in the behavior of individual participants (Bodin & Prell, 2011; Hanneman & Riddle, 2005). An ego network is a part of a social network that consists of a focal actor, all of his relations to other people, and the relationships among these people (Wasserman & Faust, 1994). The ego network characteristics, including the size of the ego network or the strength of the ties, may reflect students’ help seeking activities and their participation in problem solving (Gourash, 1978; Liccardi et al., 2007). The size of ego networks, referring to the number of network members, has a direct impact on the interaction among its members (Zhu, 2006). A larger size suggests more diverse interaction among network members (Fahy, Crawford, & Alty, 2001). The average strength of ego network ties reflects the responsiveness and attentiveness of members to each other (Fahy et al., 2001). Higher average strength of ego network ties means stronger connections among participants, which represents more responsive or emotional support (e.g., giving comments) or emotional support (e.g., showing likes). The combination of characteristics of ego networks exhibited by students in their problem solving activities represents students’ ego network patterns of problem solving. Considering social network analysis has the potential of understanding students’ patterns of participating in problem solving activities in an online discussion forum (Rabany et al., 2012), analyzing students’ ego network patterns of problem solving activities (e.g., subscribing to others’ questions, commenting on others’ questions, viewing comments made by others, and liking comments made by others) can be helpful in understanding students’ participation in problem solving activities in an online discussion forum. Additionally, given the importance of the network size and strength of ego networks, it is suggested that the facility of participation in academic problem solving using online discussion forums should consider support for the growth of ego networks or enhancing the strength of the network ties.
Based on the aforementioned concerns, designing an online discussion forum system for help seeking should consider the students’ difficulties in seeking potential helpers, and should also deliberate on facilities that stimulate meaningful interaction among students for academic problem solving. To this end, this study develops a proactive online discussion forum that strengthens online social networks in an online discussion forum. The proactive online discussion forum allowed individual learners to proactively send messages to friends or potential experts to invite advice for solving problems. This study further explored the effects of the proactive online discussion forum on students’ participation in problem solving and perception of help seeking in terms of students’ behaviors, ego network patterns, and attitude toward help seeking. The research questions in this study include:

- What are the effects of the proactive online discussion forum on students’ participation in problem solving?
- What are the effects of the proactive online discussion forum on students’ ego network patterns of problem solving?
- What are the effects of the proactive online discussion forum on students’ perception of help seeking for problem solving?

Method

Participants and the learning task

This study included 37 Vietnamese graduate students (15 males and 22 females) with basic level of English proficiency in a university in Taiwan. All participants were randomly assigned to either the experimental or control group. In all, 19 participants (8 males and 11 females) in the experimental group were asked to study material about Basic English grammar on a proactive online discussion forum where they could subscribe and respond to the questions that were raised by the other students. This group of students could also send invitation emails to their friends or experts for discussion of the inviters’ questions. On the other hand, 18 participants (7 males and 11 females) in the control group studied the same material on another online discussion forum, which had the same features as the experimental group’s proactive online discussion forum except for the function of sending invitation emails to students’ friends or to the experts recommended by the forum. The material was written in English and presented in the format of PowerPoint slides. All participants were free to ask their questions in English or Vietnamese regarding the contents of the slides on their respective online help forums and expected responses from other participants to the posted questions.

The proactive online discussion forum

As shown in Figure 1, the proactive online discussion forum consists of three main parts: question categories, discussion, and question menus. The top of Figure 1 shows three different categories of questions: self-asked, unsubscribed, and subscribed. The questions asked by a student him/herself are referred to as self-asked questions. Questions asked by other students were initially categorized as unsubscribed. Once a student subscribed to an unsubscribed question, it was then referred to as a subscribed question. To the left of Figure 1 is the area for the discussion of questions. This area includes a question, a PowerPoint slide, and a set of comments. Each question asked by a student was associated with one PowerPoint slide. As shown in Figure 1, a student marked on a PowerPoint slide and asked the following question: “In the yellow block, can I use ‘when’ instead of ‘as’?” To the right of Figure 1 is the menu that allowed the students to subscribe to or unsubscribe from a question. When a student subscribed to a question, the online discussion forum system continued sending notification emails to the student to inform him/her of the arrival of new comments on that question.

The proactive online discussion forum strengthened online social networks connecting with friends and recommended experts. With the online social networks, students were free to send invitation emails to their friends or potential experts. An invitation email consists of an inviter, the content of a question, the author of the question, and the role of the invitee. As shown in Figure 2, a student posted a question and invited friends to discuss the question. In this case, the inviter is the author of the question, and the role of the invitees is friend. The student had to add friends into the networking structure before being able to invite them to the discussion. Other members of the online help forum who received invitation emails could decide whether to subscribe to the question or merely leave it unsubscribed. On the other hand, the forum recommended potential experts to the author of a question based on members’ expertise regarding the PowerPoint slides. The members’ expertise was calculated based on peer assessment of members’ comments on questions. For example, one student received many likes on his/her comments to questions associated with a PowerPoint slide; therefore, the forum was more likely to recommend that student as a potential expert for solving problems related to that particular PowerPoint
slide. Students could also send invitation emails to the recommended potential experts. In such a case, the inviter is the author of the question, and the role of the invitees is expert.

Figure 1. A proactive online discussion forum

Figure 2. Inviting friends to participate in problem solving
Research design and procedure

Two versions of online discussion forums corresponding to their respective functions were employed in the experiment: a proactive online discussion forum and a conventional online discussion forum. The proactive online discussion forum designed for the experimental group students included basic online discussion forum features and invitation functions. The basic features, as shown in Figure 1, allowed the students to post questions, comment on questions, view the comments, and subscribe to interesting questions. The invitation function, as shown in Figure 2, allowed students to receive recommendations of potential experts and invite their friends or the recommended experts to the discussion of their questions. The conventional online discussion forum, which was designed for the control group students, had the exact same basic online discussion forum features as the proactive online discussion forum except for the invitation functions. This version mirrored the help-seeking features of typical online discussion forums that allowed students to request help in public and wait passively for other students’ responses. The procedure for the study mainly consisted of a one-hour training session and three weeks of study sessions for all participants. In the training session, the respective versions of the online discussion forums were introduced to the participants. An instructor presented the system to the students and encouraged them to use the forums to post their questions or provide comments to the questioners. In the study session, the two groups of participants were asked to study the material related to Basic English grammar. They were free to use the functions of their respective online discussion forums to resolve their questions in a convenient location and at a convenient time during the three weeks. During this session, all participants’ activities of asking questions, making comments, viewing comments, giving likes, and subscribing to questions were collected. The invitation emails sent by the experimental group were also collected for analysis. Finally, all participants were asked to complete a questionnaire that investigated their perceptions of help seeking using the respective online discussion forum.

Instruments, data collection and analyses

The instruments employed in this study included two different versions of online discussion forums and the questionnaire for investigating students’ perceptions of help seeking. There are three major variables involved in this study: students’ participation in problem solving, students’ ego networks of problem solving, and the students’ perceptions of help seeking. In both versions of the forum, the students’ participation in help seeking for problem solving was explored by analyzing their problem solving activities during their study of the material. The activities were identified according to some common behaviors of participating in an online discussion forum such as posting questions, making comments on questions, viewing comments, and subscribing to questions. All the aforementioned activities were logged by the two versions of the forum, and the frequencies of the activities were computed to compare the differences between the two forums.

The ego network of problem solving was employed to analyze the degree to which a student utilized or created social connections with other students when subscribing to questions, commenting on questions, viewing comments, or giving likes to other students’ comments. Among the properties of these ego networks of problem solving, their size and strength were computed to compare the differences between the two discussion forums. For example, to compute the size and strength of the ego networks of subscribing to questions, connections that link the subscribers with the authors of the subscribed questions were first identified. The size of the ego networks of subscribing to questions refers to the number of connections that link a subscriber with distinct authors of subscribed questions. The strength of the ego networks of subscribing to questions refers to the ratio of the number of total connections to the size of the ego networks of subscribing to questions.

To measure the participants’ perceptions of help seeking, a revised version of the help-seeking scales questionnaire (Karabenick, 2003) was adopted. Because the original items and scales were in English, the back translation process (Brislin, 1970) was performed to ensure the construct validity of the translated items and scales in Chinese. There was no disconcerting discrepancy between any of the original and back-translated item pairs used in this study. To address the construct validity of help-seeking scales in this study, Partial Least Squares (PLS), which could model latent constructs under conditions of non-normality and small sample size, was adopted with using VisualPLS (Chin, 1998). The results (Table 1) support the convergent validity of the scales: loadings ≥ 0.7; composite reliability ≥ 0.7; average variance extracted (AVE) ≥ 0.5 (Fornell & Larcker, 1981). To ensure their discriminant validity, the results (Table 2) confirm that the intercorrelations of the construct do not exceed the square root of the AVE (Fornell & Larcker, 1981), which supports discriminant validity. The questionnaire focused on variables included in most models of help seeking (e.g., Newman, 1990). Because students’ goals of and attitudes towards help seeking have been suggested as being useful for measuring help seeking, in this study, as shown in Table 3, four relevant dimensions of the students’ help-seeking were
employed: instrumental help seeking, executive help seeking, help-seeking threat, and help-seeking avoidance. The questionnaire was scored on a 5-point Likert scale, from 1 (strongly disagree) to 5 (strongly agree). The coefficient of Cronbach’s $\alpha$ was adopted to reveal that the reliability of the four scales was .72, .77, .72 and .68, respectively. The instrumental help seeking, executive help seeking, help-seeking threat, and help-seeking avoidance scales are comprised of two, two, three, and three items, respectively. The differences in the two groups of students’ perspectives on help seeking for the different online discussion forums were compared. Because the major variables involved in this study were non-normally distributed, nonparametric Mann-Whitney U tests were employed to examine differences between two groups.

Table 1. Convergent validity

<table>
<thead>
<tr>
<th>Dimensions/ items</th>
<th>Factor loadings</th>
<th>AVE</th>
<th>CR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrumental help seeking</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IN1</td>
<td>0.901</td>
<td>0.813</td>
<td>0.897</td>
</tr>
<tr>
<td>IN2</td>
<td>0.902</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Executive help seeking</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EX1</td>
<td>0.900</td>
<td>0.809</td>
<td>0.895</td>
</tr>
<tr>
<td>EX2</td>
<td>0.899</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Help-seeking threat</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TH1</td>
<td>0.725</td>
<td>0.644</td>
<td>0.844</td>
</tr>
<tr>
<td>TH2</td>
<td>0.814</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TH3</td>
<td>0.863</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Help-seeking avoidance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AV1</td>
<td>0.861</td>
<td>0.617</td>
<td>0.828</td>
</tr>
<tr>
<td>AV2</td>
<td>0.718</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AV3</td>
<td>0.770</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Discriminant validity

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Instrumental help seeking</td>
<td>0.902</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Executive help seeking</td>
<td>0.010</td>
<td>0.899</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Help-seeking threat</td>
<td>-0.0247</td>
<td>0.058</td>
<td>0.802</td>
<td></td>
</tr>
<tr>
<td>4. Help-seeking avoidance</td>
<td>-0.179</td>
<td>0.352</td>
<td>0.550</td>
<td>0.785</td>
</tr>
</tbody>
</table>

Note. The main diagonal shows the square root of AVE.

Table 3. Dimensions of help-seeking scale

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrumental help seeking</td>
<td>The extent that help sought would serve instrumental goals.</td>
<td>If I were having trouble understanding the material in this PowerPoint slide course I would ask someone who could help me understand the general ideas.</td>
</tr>
<tr>
<td>Executive help seeking</td>
<td>The extent that help sought would serve executive goals.</td>
<td>The purpose of asking somebody for help in this PowerPoint slide course would be to succeed without having to work as hard.</td>
</tr>
<tr>
<td>Help-seeking threat</td>
<td>How threatened seekers are to seek help.</td>
<td>I would feel like a failure if I needed help in this PowerPoint slide course.</td>
</tr>
<tr>
<td>Help-seeking avoidance</td>
<td>The extent to which seekers decide not to seek help.</td>
<td>I would rather do worse on an assignment I could not finish than ask for help.</td>
</tr>
</tbody>
</table>

Results

The difference in student academic problem solving

A series of nonparametric Mann-Whitney U tests was used to examine the difference between the experimental and control groups in terms of their participation in problem solving using the online discussion forums, as shown in Table 4. The results revealed significant differences in the participation of questioning, subscribing to, commenting on questions, and viewing comments on questions. The students in the experimental group had significantly higher frequencies than those in the control group for questioning ($U = 101, p = .03$) and subscribing to questions ($U = 0, p < .001$). These results indicate that the proactive online discussion forum could prompt the students to post questions and subscribe to questions of interest. In terms of commenting on questions and viewing the comments on questions, the experimental group made significantly more comments
on other students’ questions \((U = 106.5, \ p = .049)\) than the control group did. The experimental group students also had significantly higher frequencies of viewing the comments on self-asked \((U = 80, \ p = .005)\) and others students’ questions \((U = 0, \ p < .001)\). These results suggest that the proactive online discussion forum could further facilitate students’ comments on other students’ questions, and could encourage the students to view the comments relevant to their own and others’ questions.

### Table 4. Comparison of the academic problem solving of the experimental and control groups

<table>
<thead>
<tr>
<th>Activity</th>
<th>Experimental</th>
<th>Control</th>
<th>(U)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questioning</td>
<td>(M = 5.26, \ SD = 1.79)</td>
<td>(M = 4.06, \ SD = 1.16)</td>
<td>101*</td>
</tr>
<tr>
<td>Subscribing</td>
<td>(M = 61.84, \ SD = 11.97)</td>
<td>(M = 20.94, \ SD = 5.32)</td>
<td>0***</td>
</tr>
<tr>
<td>Commenting on</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-asked</td>
<td>(M = 4.16, \ SD = 2.73)</td>
<td>(M = 2.98, \ SD = 1.75)</td>
<td>126.5**</td>
</tr>
<tr>
<td>Other students’</td>
<td>(M = 22.89, \ SD = 8.49)</td>
<td>(M = 17.44, \ SD = 6.47)</td>
<td>106.5*</td>
</tr>
<tr>
<td>Viewing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-asked</td>
<td>(M = 13.53, \ SD = 8.57)</td>
<td>(M = 6.84, \ SD = 5.44)</td>
<td>80**</td>
</tr>
<tr>
<td>Other students’</td>
<td>(M = 96.58, \ SD = 44.62)</td>
<td>(M = 46.17, \ SD = 15.21)</td>
<td>0***</td>
</tr>
</tbody>
</table>

*Note. \(n = 18, \ ^*n = 19; \ p < .05; \ ^{*}{*} p < .01; \ ^{***}p < .001.\)

### The difference in student ego networks

A series of nonparametric Mann-Whitney \(U\) tests were used to examine the difference between the experimental and control groups in terms of the size and strength of students’ ego networks of problem solving. These ego networks for analysis were identified according to the students’ interaction with other students including subscribing to others’ questions, commenting on others’ questions, viewing comments made by others, and liking comments made by others. For example, the size of the ego network of subscribing represents the number of distinct members that a student subscribed, and the strength of the ego network of subscribing represents the average number of questions that a student subscribed to each member. As shown in Table 5, the students in the experimental group had significantly smaller size \((U = 79, \ p = .005)\) but higher average strength \((U = 0, \ p < .001)\) ego networks of subscribing than the control group did. This result suggests that the proactive online discussion forum could facilitate student concentration on fewer members of the forum when subscribing to questions, and could also enhance students’ attention to interested questioners. With regard to giving comments on questions, the experimental group students had higher average strength \((U = 83, \ p = .007)\) of their ego networks of commenting than the control group did. This result suggests that the proactive online discussion forum may help students offer academic support that is more responsive to the requests made by members of the ego networks.

### Table 5. Comparison of the size and average strength of the ego networks of the experimental and control groups

<table>
<thead>
<tr>
<th>Activity</th>
<th>Experimental</th>
<th>Control</th>
<th>(U)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subscribing</td>
<td>(M = 10.16, \ SD = 3.02)</td>
<td>(M = 12.94, \ SD = 1.55)</td>
<td>79**</td>
</tr>
<tr>
<td>Commenting on</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>(M = 6.77, \ SD = 3.03)</td>
<td>(M = 1.61, \ SD = 0.31)</td>
<td>0***</td>
</tr>
<tr>
<td>Strength</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Viewing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>(M = 10.84, \ SD = 2.46)</td>
<td>(M = 10.28, \ SD = 2.49)</td>
<td>149</td>
</tr>
<tr>
<td>Strength</td>
<td>(M = 2.08, \ SD = 0.47)</td>
<td>(M = 1.66, \ SD = 0.29)</td>
<td>83**</td>
</tr>
<tr>
<td>Liking</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>(M = 13.95, \ SD = 2.64)</td>
<td>(M = 8.28, \ SD = 3.12)</td>
<td>29***</td>
</tr>
<tr>
<td>Strength</td>
<td>(M = 7.11, \ SD = 3.48)</td>
<td>(M = 6.12, \ SD = 2.36)</td>
<td>147.5</td>
</tr>
</tbody>
</table>

*Note. \(n = 18, \ ^*n = 19; \ p < .05; \ ^{*}{*} p < .01; \ ^{***}p < .001.\)

However, in terms of the ego network of viewing, that of the students in the experimental group was significantly larger \((U = 29, \ p < .001)\) than that of the control group. This result may reveal that the proactive online discussion forum could also assist students in extending their social networks, allowing them to view comments made by more diverse members of the forum. With regard to liking comments, the experimental group students had higher average strength \((U = 98, \ p = .026)\) in the ego network of liking than the control group.
did. This result suggests that proactive online discussion forum may lead to social networks that are more attentive to offering emotional support for the members of the ego networks.

The difference in student perceptions of help seeking

A series of nonparametric Mann-Whitney U tests was used to examine the difference between the experimental and control groups in terms of their perceptions of help seeking using different online discussion forums. As shown in Table 6, the results obtained by Mann-Whitney U tests revealed significant differences in the scale of instrumental help seeking ($U = 104.5, p = .036$), and the scale of help-seeking threat ($U = 93, p = .014$). The students in the experimental group had higher scores of instrumental help seeking in solving problems using online discussion forums than did those in the control group. These results suggest that the proactive online discussion forum could improve the students’ tendencies to seek instrumental help for solving academic problems. In addition, the results reveal that the students in the experimental group had lower scores for their help-seeking threat of seeking help using the online discussion forum than did those in the control group. This result may imply that the proactive online discussion forum could reduce the threat of help seeking in solving problems.

<table>
<thead>
<tr>
<th>Help-seeking scales</th>
<th>Experimental</th>
<th>Control</th>
<th>U</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>Instrumental help seeking</td>
<td>4.06</td>
<td>0.68</td>
<td>3.67</td>
<td>0.51</td>
</tr>
<tr>
<td>Executive help seeking</td>
<td>2</td>
<td>0.71</td>
<td>2.22</td>
<td>0.94</td>
</tr>
<tr>
<td>Help-seeking threat</td>
<td>2.18</td>
<td>0.39</td>
<td>2.54</td>
<td>0.46</td>
</tr>
<tr>
<td>Help-seeking avoidance</td>
<td>2.63</td>
<td>0.46</td>
<td>2.85</td>
<td>0.53</td>
</tr>
</tbody>
</table>

Table 6. Comparison of the experimental and control groups’ help-seeking scales

Note. *n = 19, †n = 18; * p < .05.

Discussion and conclusion

This study developed a proactive online discussion forum supporting help-seeking activities for students’ problem solving, and further assessed its effectiveness by exploring students’ participation in problem solving and their perceptions of help seeking. It was found that the students in the experimental group significantly outperformed those in the control group in terms of participation in problem solving. The overall results revealed that the proactive online discussion forum supported by online social networks could promote students’ participation in help seeking for problem solving when compared with the conventional online discussion forum. These findings are consistent with the previous suggestions (Bull et al., 2001; Melrose, 2006) that providing students with social supports in an online community would be helpful for the interaction between help seekers and potential helpers.

Many technological and pedagogical issues of academic problem solving using online discussion forums were addressed in this study. In terms of the technological aspect, this study strengthened online social networks that associate students with their friends and system-recommended experts to encourage invitation to potential helpers. This feature, which strengthens the social connections among members of the online discussion forum, provided the students with reliable help sources for help seeking and problem solving. This facilitates help-seeking activities by proactively engaging potential helpers in solving problems rather than passively waiting for responses from other members. Unlike other facilitation techniques that require particular facilitators such as teachers or board managers to keep the discussion on track or to promote engagement in the discussion (Hew & Cheung, 2008), the proactive online discussion forum encourages help seekers and invited helpers to play active roles of facilitating the discussion or participating in solving problems. Such active roles in help seeking activities may assist students in fostering positive help seeking strategies. Such positive help seeking strategies include adaptive help seeking strategies, which help seekers show an awareness of the difficulties they cannot overcome on their own, and exhibit self-determination to remedy those difficulties by requesting assistance from more knowledgeable individuals for independent learning (Newman, 2002). Consequently, an online discussion forum incorporated with online social networks enabled a proactive help-seeking environment that supports the students in problem solving while considering the students’ concern about seeking potential helpers. The results showed that the students who utilized the proactive online discussion forum posted and subscribed to more questions than did the students who utilized the conventional online discussion forum. The students also demonstrated more comments on questions and more viewing of comments than did the control group students. These improvements in participation of problem solving confirmed the effectiveness of the online social network.
approach in terms of reducing students’ difficulties in seeking or giving help in an online discussion forum. The findings of this study provide some evidence that the online social network approach to the design of online discussion forums is beneficial for students’ help seeking for problem solving.

The difference in the ego network patterns of help seeking for problem solving also evidenced the influence of the online social network approach on social participation in an online discussion forum. Some novelty features provided by the proactive online discussion forum may have influenced the students’ interactions with other members of the forum. For example, the forum allowed students to explicitly invite their friends to the discussion of questions. With the invitation, the inviters can anticipate feedback from their friends, and their friends may then feel responsible for responding to the questions. Therefore, the online social network approach that reinforces social connections may influence the patterns of ego networks of problem solving. Taking the ego network of subscribing to questions as an example, the results show that the explicit invitation to friends or experts may lead to smaller sized ego networks of subscribing. The invitation feature helps students concentrate on fewer members of the online discussion forum when subscribing to questions. The results also show that explicit invitations to friends or experts may assist students in demonstrating higher strength of ego networks of problem solving, which are more attentive to offering academic assistance (e.g., comments) and emotional support (e.g., likes) to the help seekers in their social networks. This suggests that the invitation feature may also facilitate in-depth participation of help givers. From a cooperative learning perspective, students’ strengthened social networks may reflect a cooperative group that could improve productivity, personal responsibility, and members’ willingness to complete difficult tasks (Johnson & Johnson, 1999). Therefore, the formation of such attentive social networks seems especially important in motivating help seekers and help givers to participate in problem solving in online discussion forums because the help givers in the attentive social networks are responsive to help seekers’ requests. In this sense, techniques or strategies that help students form a group in which members have close relationships (e.g., friendship) with each other may play positive roles in promoting social participation of help seeking and help giving in an online discussion forum.

The increase in the students’ tendencies of seeking instrumental help, and the decrease in the students’ help-seeking threat also evidenced the effects of the proactive online discussion forum. Some technical features adopted by the forum may have enhanced the tendency of instrumental help-seeking and reduced the help-seeking threat in seeking help for solving academic problems. For example, familiar friends as target sources of help seeking may influence attitudes towards seeking help. According to Newman (2000), where there was greater familiarity and friendship among students, help seeking was more likely to be successful and hence more likely to be positively reinforced as a useful strategy. Students’ help-seeking threat may be reduced by the familiarity of help givers, and students’ value of instrumental help seeking could be enhanced during the successful help seeking process. In addition to the invitation to their friends, students also sent invitation emails to the experts recommended by the system. A Mann-Whitney U test result shows that the students sent more invitation emails to their friends than to the system-recommended experts (8.72 vs. 5.76 per question). This preference may suggest that the friend relationship had a positive impact on stimulating help seeking in the online discussion forum when compared with the recommendation of potential experts. However, the result of Pearson’s correlation revealed that the frequency of inviting experts for discussion was negatively correlated with help-seeking threat in solving problems ($r = -.47, p = .044$). This may suggest that introducing competent experts into the help seeking process for solving academic problems tends to have a positive effect on reducing help-seeking threat. Therefore, the invitation to friends and potential experts should be employed complementarily to help students obtain friendly and credible comments as well as to assist them in developing different support groups. For example, a novice member may initially be suggested to invite system-recommended experts to their questions and gradually develop friendships with the potential experts or other members attracted by the discussion.

A series of comparative results revealed that the students exhibited more frequent and in-depth participation when they utilized the proactive online discussion forum to resolve their problems. Given that problem solving is viewed as a powerful way of promoting knowledge construction (Anderson, 1993), further application of the proactive online discussion forum could be carefully reframed as an educational tool or environment from pedagogical perspectives to promote students’ learning or help-seeking strategies through problem solving. For example, the proactive online discussion forum could include assignments containing problems relevant to the curriculum to guide students to learn. Then a teacher could encourage the students to discuss the problems in the proactive online discussion. Additionally, students’ behaviors of seeking help for problem solving (e.g., inviting potential helpers or giving comments) or the structures of ego networks (e.g., size or strength) could provide a conceptual framework for assessing students’ help-seeking strategies or for predicting interaction patterns among students. Finally, the application of the proactive online discussion forum to learning and instruction in some specific subjects, such as computer programming, which often require students’ understanding and use of
abstract concepts to solve problems, should be carefully examined. Nevertheless, the lack of a significant difference between the two groups on the executive help-seeking and the avoidance of help-seeking subscales may be the result of the small-scale sample or other personal factors such as goal orientation (Karabenick, 2003). Hence, further work needs to be undertaken with a larger sample and considering more personal factors to provide more robust evidence.

Acknowledgements

This research was partially funded by Ministry of Science and Technology, Taiwan, under grant number MOST 106-2511-S-155-004-.

References


Analyzing Group Dynamics of a Digital Game-based Adventure Education Course

Chang-Hsin Lin and Ju-Ling Shih*

Department of Information and Learning Technology, National University of Tainan, Tainan, Taiwan // changhsinlin1206@gmail.com // juling@mail.nutan.edu.tw

*Corresponding author

(Submitted November 12, 2016; Revised February 23, 2017; Accepted June 30, 2017)

ABSTRACT

This study reported the group development training results using a digital game-based adventure education course. In the research, six traditional adventure education activities were developed into six digital games that are assembled and practiced in the face-to-face course. The course was designed based on Tuckman’s team development model in which students are trained to have positive human interactions and work as a team. With Hill Interaction Matrix (HIM) and focus group, the processes of students’ oral interactions in the course were documented and analysed to see each groups’ group dynamics and group development conditions. The results showed that groups of all kinds can be developed into effective groups and had positive human interactions after the course.

Keywords
Adventure education, Digital game-based learning, Team development, Group dynamic, Hill Interactive Matrix

Introduction

Group dynamic is the power while group members are interacting. It can keep the group development and team-building process running, and influence behaviours of each group member and function of groups (Cartwright, 1951). Training for team-building has many types of methods. Adventure education has been an effective way for team-building. Participants are normally divided into groups to work with others, familiar or not, to become cooperative partners. Reflections are conducted after every activity led by the facilitators. Through the reflections, participants retrospect the process to train their skills of problem-solving or team work, and learn the lessons of each activity that they can further use in daily lives (Cooley, Burns, & Cumming, 2015). Nevertheless, some traditional activities were difficult to implement due to rules which are hard for the facilitators to manage, complicated activity equipment to prepare, and occasional bad weather. However, few studies of any kind have examined the methods to solve the problems. It is also rare to find adventure education to be conducted in the any degrees of digital forms.

Therefore, this research hopes to provide a new effective adventure education course with the implementation of digital games that are easy to conduct, and can increase group interactions. Since teenagers in this digital era are generally interested in digital games, it might be a great opportunity to introduce digital technology into the activities. The digital games would make them immerse in the gaming situation, enhance learning motivation, increase learning effectiveness, and encourage them to express thoughts and emotions (Batson & Feinberg, 2006; Przybylski, Ryan, & Rigby, 2009; Schell, 2014).

Therefore, this paper aims to integrate digital games into the adventure education course to design a digital game-based adventure education course. Six traditional activities were chosen to be developed into digital games. The digital games in the course were built by Unity3D. With the advantages of cross-platform and high simulations, the six traditional activities were developed into two network computer games, two tablet games and two motion-sensing games. Those six games were used in the five stages according to Tuckman’s team development model including forming, storming, norming, performing, and adjusting stages. In the course, students can play digital games, interact with their members, learn positive human interactions, and go through the team-building process in every stage. After the course, groups are expected to become effective groups. Hill interaction matrix and focus group interviews were used to investigate the effectiveness of the course. Thus, in this paper, the effects of digital game-based adventure education course on learning and human interactions would be examined. Specific research questions to be addressed are as follows: How were the group dynamics and the group development processes of each group in the digital game-based adventure education course? How did group dynamics in the reflection sessions influence group developments?
Literature review

Adventure education

Adventure education is a series of adventurous and challenging activities that allows students to participate and experience group works. After each activity, the facilitator would guide students through the reflection process, allow them to internalize the theme, and further apply them into real life practices (D’Amato & Krasny, 2011; Ewert, Sibthorp, & Sibthorp, 2014; Lund, & Tannehill, 2010). In adventure education courses, teachers act as facilitators to allow students to explore and experience course content, and immerse in learning by doing exercises (Howden, 2012). In such a way, learning has become more interesting and interactive, and learning concepts would be an activity that is less static, which can enhance learning effectiveness (Sutherland & Stuhr, 2014).

So far, adventure education incorporates outdoor activities or portable adventure activities that involve physical activities (Battey & Ebbeck, 2013; D’Amato, & Krasny, 2011). The activities are mostly conducted in groups to train group members for group coherence, leadership, communication skills, and efficiency. The activities of adventure education can be divided into three stages (Henton, 1996; Lin, Shih, & Hsu, 2013): (1) Brief: Before activities begin, facilitators have to talk to students about activity rules, goals, situations, and restriction. It is to call for students’ attention to rules and goals; (2) Activity (Game): Students start to experience activities. Facilitators play the role of assistants who help students finish the courses. They are not lecturers in class; (3) Debrief (Reflection): Facilitators guide the students to understand the meanings of the activities through reflections. Nevertheless, the group development process and member relationships in the adventure education courses are strengthened but hardly observed. Tuckman (1977) has given small group development a clear definition through five stages, namely forming, storming, norming, performing, and adjusting, so that groups can become effective groups (Tuckman & Jensen, 1977). Until now, Tuckman’s theory is widely used in various fields for team-building (Bonebright, 2010; Garfield & Dennis, 2012; Haines, 2014), especially in the business organizations and industries. It provides a framework for the observation and understanding of the process of team-building.

Sometimes, facilitators encounter difficulties in adapting to weather, environment, and role managements when conducting traditional activities. Thus, this research chooses adventure education activities that have high requirements for physical environment setups, and attempts to make those activities more feasible in common practices.

Digital game-based adventure education course

The only one study regarding adventure education using digital game was Hsu and Shih (2013) in which they found that participants can learn that abilities of problem-solving and positive cooperation through digital game as well as in the physical form. It proved that adventure education activities can be effectively practiced in the form of digital games.

Digital game-based learning (DGBL) has earned its fame in the recent years. Its original purpose is to use digital games to increase learners’ learning motivation, and further to improve their learning abilities and effectiveness. Many research have obtained positive evaluation results in various fields and subjects (Hong, Cheng, Hwang, Lee, & Chang, 2009; Hsiao, Chang, Lin, & Hu, 2014; Suh, Kim, & Kim, 2010).

Adventure education, as closely related to counselling, is normally performed through face-to-face mode which imposes tacit pressure to the participants. To relieve the inter-personal tensions and bridge the communication between parties, participants need to have strong human interaction skills, such as listening and empathy, to lower down each other’s mental defence. Game-based counselling (GBC) was first appeared in Hsu and Shih’s (2012) study suggesting using digital games in the counselling process can enhance counselling effectiveness. Further down the road, when digital games alone can generate positive and similar effects as traditional counselling sessions, they can be called as Digital Counselling Games (DCG).

This study, slightly different from GBC and DCG, adds and emphasizes on the reflection session conducted by the facilitator after each game, which guides the participants to discuss through a series of targeted issues. The participants retrospect their gaming process, think about the implications of the games, and internalize the group development themes, such as interaction, communication, and cooperation; and work through the process to become effective groups with their group members. In order to understand the interaction process, Hill
Interaction Matrix is used to observe the oral interactions in the reflection session to evaluate the effectiveness of group development of the course.

**Hill Interaction Matrix**

Hill Interaction Matrix (HIM) is an oral interaction analysis system developed by Hill (1971) which is appropriate for investigating the interaction and change of group therapy. The analysis is done through multi-dimensional systematic categorization. Oral interactions include two major factors: Content and Work factors (Barlow, 2013; Berg, Landreth, & Fall, 2013). Content factors include Topic (I), Group (II), Personal (III), Relationship (IV) dimensions. Topic (I) refers to all general group development issues. Group (II) includes issues such as activity execution, leadership, and group cooperation; that generally come from individual feedbacks and their doubts to groups. Personal (III) includes issues concerning specific member, such as his personality and behaviours, etc. Relationship (IV) refers to issues concerning members’ interactions and inter-relationships with the group members.

Work factors define types of oral interactions, which include Responsive (A), Conventional (B), Assertive (C), Speculative (D), and Confrontative (E). Responsive (A) dialogue refers to communications that are intrigued by the facilitators; members only respond with short answers without having basic social communication skills. It is often carried out by mental disabilities. Conventional (B) dialogue refers to greetings, chats, entertaining gossips without having much impact to group development. Assertive (C) dialogue refers to messages about self, mostly on defensive, divergent, replenish, and even assault words. Speculative (D) dialogue refers to discussions that are rational and supportive for members’ thinking, observation, and comprehension. Confrontative (E) communication refers to the interactions that the group leader or a member tends to force the important person to face the avoided problems, which might lead to risks and intensity.

HIM result presentation uses horizontal axis to show the content factors, and vertical axis to show work factors, which forms a 4x5 matrix, with total of 20 categories (Table 1). From the analysis of oral interactions using the matrix, the facilitators can observe whether the participants’ discussion dialogues have therapeutic functions.

**Table 1. Hill interaction matrix**

<table>
<thead>
<tr>
<th>Work/Content</th>
<th>Topic(I)</th>
<th>Group(II)</th>
<th>Personal(III)</th>
<th>Relationship(IV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responsive (A)</td>
<td>I A</td>
<td>II A</td>
<td>III A</td>
<td>IV A</td>
</tr>
<tr>
<td>Conventional (B)</td>
<td>I B</td>
<td>II B</td>
<td>III B</td>
<td>IV B</td>
</tr>
<tr>
<td>Assertive (C)</td>
<td>I C</td>
<td>II C</td>
<td>III C</td>
<td>IV C</td>
</tr>
<tr>
<td>Speculative (D)</td>
<td>I D</td>
<td>II D</td>
<td>III D</td>
<td>IV D</td>
</tr>
<tr>
<td>Confrontative (E)</td>
<td>I E</td>
<td>II E</td>
<td>III E</td>
<td>IV E</td>
</tr>
</tbody>
</table>

HIM analysis method was used in many studies in counselling and medical fields to explore more in-depth psychological investigations (Kivlighan, 2014; Latour & Cappeliez, 1994; Pan, Deng, Fan, & Yuan, 2012). Consequently, this research used HIM to analyse the group interaction process of the adventure education course to investigate the group development conditions.

**Course and game design**

Tuckman used his group development model to explain the process of groups while becoming effective ones, which include five stages namely, forming, storming, norming, performing, and adjusting (Tuckman & Jensen, 1977). The stages are sequential, and every stage is essential for team building. The failure of one stage would cause difficulties for the groups to move onto the following stages (Johnson, Suriya, Yoon, Berrett, & La Fleur, 2002). In order to ensure that the adventure education course designed in this research would reach the goal of successful team-building, six adventure education activities were chosen from the book, One Hundred and Fifty Suggested Experiential Education Activities (Hsieh, Wang, & Chuang, 2008), and transformed them into digital games. They were Polar Bear and Hole, Cooperative Puzzle, Chessboard Maze, Moon Ball, Group Balance, and Calculator. The six activities are those that are difficult to carry out in the physical environment due to some specific requirements, or uncontrollable factors such as weather or physical space. As a result, the activities were made into digital adventure education games based on their original goals, were designed with related counselling theories and were presented in suitable digital forms. The six digital games were verified by two adventure education experts and were in accordance with their original activities. They were unlike commercial
games that were designed for general themes, they had specific training goals. After each game, one reflection session was performed to allow the participants to understand the implied meanings of the games.

Thirty students were divided into groups of five. Every two groups were invited to participate the course on the same day. The whole course was repeated three times/days to allow all six groups to complete the course. The instructional design is done as follows.

**Forming stage**

Members are unfamiliar with each other in the beginning stage, and they try to search for group goals together. Members are searching for their roles in the group at this time. Using ice-breaking activities would give members chances to communicate, share, present their opinions, and get familiar with group dynamics, common values, and individual thoughts. It allows the members to build relationships between them. The game used in this stage is Polar Bear and Hole. With a lead story, participants try to solve the riddle with reasoning skills and speculations. The main theme of the game, Polar Bear and Hole, is to listen. Participants learn to listen to embedded meanings in the communications. “Listen” is the first ability the participants need to strengthen in the interpersonal interactions.

**Storming stage**

Members start to find their own goals, have their own thoughts, and manage things from their own perspectives at this stage. Individuals start to make influences to the group. In this stage, activities that require little oral communications were chosen so that all members observe each other’s needs and create individual values during the game. At this point, members have internal dialogues, face group challenges, make improvements, and move toward effective group. The game used in this stage is Cooperative Puzzle. Every member gets three puzzle pieces randomly out of 15 pieces. Each has to complete a square puzzle of his own. Five members have to complete five square puzzles at the same time in order to finish the game. In the process, members have to exchange puzzle pieces without any oral communication. Also, puzzle pieces can only be given instead of asked. The goal of the game is to make the members observe others’ needs to reach the group goal, think about what they can do for the group, and identify their roles in the group. They have to learn that individual accomplishment is not group accomplishment. Every member needs to work toward group welfare to reach final success. Another game used in this stage is Chessboard Maze. Members have to take turns to pass a 4 by 5 chessboard maze which is full of landmines. There is only one safe route. Once stepped the landmine, the player has to return to the start point by going back on the same route. The members take turns to try out the maze until finding the correct route. What other members have to do at the same time is to figure out and memorize the correct route, and then give directions to the player in the maze. The purpose of the game is to build trust between group members, define group strategies, and learn the value of trial-and-errors. Group members have to pass their experience on to other members, and work with bravery when it seems to be dead-end.

**Norming stage**

In this stage, groups start to have group goals centered by the leader and rules. Members learn to accept others’ ideas, willing to make adjustments to changes of conditions, to redefine their positions in the group, and play their own roles faithfully in the groups. They communicate, trust each other, and use strategies to work effectively toward group goals. The game used in this stage is Moon Ball. Members have to take turns to pat the ball for the longest time they can accomplish. The purpose of the game is to train members to collaborate with each other to achieve the goals. Problem-solving strategies and team work spirits are the core abilities for this game. They are generated from the actions, but not from the insights.

**Performing stage**

In this stage, the group can see the results of their efforts. Through all the frustrations they encountered in the earlier stages, the groups should have known the strategies to complete the task and create the peak experience. Members now search for dynamic changes and further advancement. They are willing to support colleagues, and can acknowledge the common goals with individual prospects and group ideals. The game used in this stage is Group Balance. Three members play the game in each session; one on the virtual balance board keeping balance
and the other two members stand at the two sides controlling the balance and heights of the board. They have to push the board upward with equal strength until the board reaches the top while keeping the person in the middle stand steadily on the board. To complete the mission, members should trust each other, and find collaboration strategies. The balance of the board signified the agreements between members; however, the disagreements are the group dynamic that move the group forward. Positive arguments would stimulate motivations, but negative arguments would hinder group’s creativity.

**Adjusting stage**

Group development progress would come to the end as members share their experiences and create common memories. After this point, groups might face member changes; when new members join the group, the group development would enter another cycle. The game used in this stage is Calculator. In this game, there are 30 ladybugs in the virtual garden with numbers 1 to 30 on their backs. Group members have to cooperatively catch the ladybugs in consecutive order in their highest speed. This exciting game is actually testing the members to re-identify their roles in the group, and work effectively as a group.

**Research methods**

This study is about group development in digital game-based adventure education course so qualitative research design was used to conduct in-depth analysis on group dynamics and group development (Figure 1). Thirty participants, aged between 19 and 25, from two different colleges volunteered to participate in the study. They were randomly divided into six groups so the group members were not familiar with each other in the beginning.

---

**Digital Game-based Adventure Education course**

<table>
<thead>
<tr>
<th>Theory</th>
<th>Interpersonal performance, Cooperative Learning, Adventure Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goals</td>
<td>Cognitive restructuring, Problem-solving, Effective group, Interpersonal interaction</td>
</tr>
<tr>
<td>Stages</td>
<td>Norming, Storming, Forming, Performing, Adjusting</td>
</tr>
<tr>
<td>Games</td>
<td>Polar Bear and Hole, Cooperative Puzzle, Chessboard Maze, Moon Ball, Group Balance, Calculator</td>
</tr>
</tbody>
</table>

**Hill Interaction Matrix**

<table>
<thead>
<tr>
<th>Work</th>
<th>Content</th>
<th>Topic(I)</th>
<th>Group(II)</th>
<th>Personal(III)</th>
<th>Relationship (IV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional(B)</td>
<td>I</td>
<td>B</td>
<td>II B</td>
<td>III B</td>
<td>IV B</td>
</tr>
<tr>
<td>Assertive(C)</td>
<td>I</td>
<td>C</td>
<td>II C</td>
<td>III C</td>
<td>IV C</td>
</tr>
<tr>
<td>Speculative(D)</td>
<td>I</td>
<td>D</td>
<td>II D</td>
<td>III D</td>
<td>IV D</td>
</tr>
<tr>
<td>Confrontive(E)</td>
<td>I</td>
<td>E</td>
<td>II E</td>
<td>III E</td>
<td>IV E</td>
</tr>
</tbody>
</table>

*Figure 1. Research structure*

The whole course was conducted in repetition for three days, six hours each. The same facilitator led 10 students of two groups through the course in each day. In six hours, each game was run for one hour including 40 minutes game time and 20 minutes group reflection session. The reflection sessions were conducted after each of the six games for the members to retrospect the gaming process and think about the themes of the games. In the course,
In order to investigate how group interactions influence team development, HIM was implemented for analysis (Hill, 1997). The dialogues in all the reflection sessions were recorded and analyzed sentence by sentence. The data were encoded into HIM quadrants by the primary researcher along with the facilitator. Since Responsive (A) is only used for mental disability patients, this research discarded the aspect, and came up with total of 16 oral categories. Two skilled coders with adventure education background were invited. Analysis revealed the reliability of the Multi-rater Fleiss Kappa coefficient $k$ to be 0.85. It achieved a high degree of consistency (Landis & Koch, 1977).

At the end of the course, focus group interviews were conducted for 20 minutes to understand the participants’ perceptions to the group interactions throughout the course, and receive overall feedbacks to the reflection sessions. The interview themes were mainly about their impressions to their group members, the interactions during the games, the relationships between members, the strategies used to complete the missions, and the difficulties they encountered during the process. The discussions of focus group interviews were also transcribed for analysis.

**Research results**

**Group dynamic analysis**

In this section, group dynamics of each of the six groups were analyzed in terms of the interaction frequencies in the four HIM quadrants. The HIM analysis results of Group 1 to Group 6 were shown in Figure 2 to Figure 7 respectively.

The results show that the interactions of all the groups in all stages laid mostly in Quadrant II which is about group issues. It is mainly because the facilitator focused more on the discussions about the gaming processes instead of about their personal feelings towards other members and games so most of the conversation issues were concerning group interactions (Quadrant II).

Looking into each individual group, we can see various group dynamics. In Group 1, two of the members were active speakers who responded to the facilitators more willingly which resulted to a more dynamic discussion of the whole group. It was comparatively easier for the facilitator to bring the group discussion from issues in Quadrant II to Quadrant IV. Since they were more willing to tackle the issues concerning group relationships (Quadrant IV), they were more effective in maintaining positive group interactions which led to better game performances.

Group 3 had slightly different curve from other groups on the storming stage in which the games Cooperative Puzzle and Chessboard Maze were conducted. The group members tend to be more introverted, that they were shy in communicating with each other. Their interaction frequency rose to the peak in Chessboard Maze when the facilitator required the members to talk. As a result, this group entered the storming stage since then.

Group 5 and 6 had less interaction in the storming stage in all four quadrants because they complete the missions in Chessboard Maze too quickly. From the interviews they said they happened to guess the answers right so they finished the game fast. From the background analysis it is found that the members in these two groups were with more digital technology experiences that they seem to handle technology more easily. Therefore, the members of the two groups did not experience the interaction conflicts and frustrations that normally exist in the storming stage. Thus, the facilitator was not able to guide the group through difficulties that were expected to happen in the group development. As a naturally consequence, the two groups spent more time on the following games to reach the same group development goals as others since they did not nurture sufficient group spirits in the storming stage.
Figure 2. Group dynamics of Group 1

Figure 3. Group dynamics of Group 2

Figure 4. Group dynamics of Group 3
Table 2 shows the sum of oral interaction frequencies of all groups in all stages in HIM classification. Related research indicated that facilitator’s guided talks and directions could greatly influence the interactions between members (Hill, 1977). In this study, members had more conversation focused on Conventional (B) (78%) and Speculative (D) (20.34%) which showed that the facilitator and the members reflected more issues about group activities. In this research, communication content were mostly focused on Group (II), Personal (III) and
I never talk to strangers before, but now I was able to through conflicts (Gilley, Morris, Waite, Coates, & Veliquette, 2014; Tuckman & Jensen, 1977). I feel more comfortable now. I found group interactions were improved, and working strategies were generated. For example, members thought that “I feel every reflection session was important and helpful to group development.” In the beginning, we feel unfamiliar with each other. But to the last game, we get to know each other’s characteristics, and know how to interact with each other (St - 01001, St - 01002).”

Crossing with the observation records, it is found that the facilitator tended to guide the reflection sessions in the style of Conventional (B) followed by Speculative (D) to treat things rationally, stimulate other members to think, conduct rational discussions to increase their comprehension abilities to problems. In terms of the topic transactions, the facilitator used issues about Group (II) followed by Relationship (IV) to discuss the member interactions in the games. After that, issues switched to Personal (III) to investigate members’ characteristics and behavioral models. Meanwhile, the reflection issues least focused were Assertive (C) (1.35%) and Confrontative (E) (0.31%). It showed that the facilitator and members made little discontentment statements. All six groups were in peaceful, respective, safe and positive environment.

<table>
<thead>
<tr>
<th>Work/Content</th>
<th>Topic(I)</th>
<th>Group(II)</th>
<th>Personal(III)</th>
<th>Relationship(IV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional (B) (78%)</td>
<td>12(0.62%)</td>
<td>1086(56.39%)</td>
<td>73(3.8%)</td>
<td>331(17.19%)</td>
</tr>
<tr>
<td>Assertive (C) (1.35%)</td>
<td>16(0.83%)</td>
<td>5(0.26%)</td>
<td>0(0%)</td>
<td>5(0.26%)</td>
</tr>
<tr>
<td>Speculative (D) (20.34%)</td>
<td>20(1%)</td>
<td>249(12.93%)</td>
<td>57(2.97%)</td>
<td>66(3.44%)</td>
</tr>
<tr>
<td>Confrontative (E) (0.31)</td>
<td>0(0%)</td>
<td>5(0.26%)</td>
<td>0(0%)</td>
<td>1(0.05%)</td>
</tr>
</tbody>
</table>

Based on those results, it is affirmative that most of the group interactions were positive. Thus, in this section, the process of group development would be analyzed.

In an overview, all six groups finished all the games in 40 minutes. Comparing the total interaction frequencies of all six groups in the reflection session of each stage, the research has reached some findings. Group 1 to 4 had great peak at the storming stage, especially at the game Chessboard Maze. The participants all pointed out in the focus group about the importance of this game to group development. “I never talk to strangers before, but now I would try to communicate with others. When the game started, we distributed works among members and try our best to achieve the goals (St-01001, St-01002).” Throughout the course, members work from being strangers to being positive and effective groups.

On the contrary, Group 5 and 6 had interaction frequency drop in the storming stage and peak in the adjusting stage. They finished the game Chessboard Maze to quickly that they did not have chance to work on the group relationships. There is little infusion between members and they missed the process of dealing with communicative frustrations. Until they reached the stage of the game Calculator, they started to have massive amount of communication since the groups need to make up the missing process of necessary group development. “In the beginning, we feel unfamiliar with each other. But to the last game, we get to know each member’s characteristics, and know how to interact with each other (St-03002, St-03005).”

Overall speaking, all six groups became effective groups after the course regardless of their interaction types through the group development stages. The only differences between the two kinds of groups are the game time they spent in the last stage for group development. Group 5 and 6 spent half an hour more in the last stage to reach the same level of group development (Table 3). The storming stage of group development is considered as an opportunity to increase positive interaction through conflicts (Gilley, Morris, Waite, Coates, & Veliquette, 2010; Pelegri Morita & Marie Burns, 2014; Tuckman & Jensen, 1977). Therefore, if groups do not have sufficient interactions in the storming stage, the speed of group developments would be hindered.

<table>
<thead>
<tr>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
<th>Group 5</th>
<th>Group 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>14 min</td>
<td>23 min</td>
<td>15 min</td>
<td>27 min</td>
<td>40 min</td>
<td>52 min</td>
</tr>
</tbody>
</table>

Participants generally agreed that the reflection sessions after each of the games were helpful for group development, and had positive influence to the following games. “I feel every reflection after the game is important and helpful to group development.” Through the guided reflections, group interactions were improved, and working strategies were generated. For example, members thought that “I found group communication started from the game Moon Balls. We started to divide works between us (St-02042).” From positive interactions and discussions, group members have more thoughts toward the group goals and better understandings to each other. For example, members thought that “Since everyone has different perspectives, the communication made us to think more and learn from the previous experiences.”

59
Correlation between group dynamic to group development

It is found that the treatment effect of the reflection session in the storming stage has great impact to group development after the storming stage (Gilley et al., 2010; Pelegrini Morita & Marie Burns, 2014; Tuckman & Jensen, 1977). The game in adjusting stage is presented to examine whether the group development is effective. One the group development is not working well, the finishing time in adjusting stage would be affected. Comparing the finishing time of Calculator in adjusting stage and the reflection content in storming stage of every group, the influence of group dynamic to the effectiveness of group development can be analyzed. To investigate the correlation between the two (Calculator’s the finishing time and the reflection content in storming stage), Pearson product-moment correlation coefficient was used (Table 4). Since the storming stage included two games, Cooperative Puzzle and Chessboard Maze, their effects, in terms of various interaction frequencies, were checked both separately and jointly with the finishing time of the adjusting stage.

The research shows high positive correlation between the finishing time of Calculator in the adjusting stage with Cooperative Puzzle Quadrant II \( (r = .856^*), \) sum of Quadrant II & IV \( (r = .847^*), \) and sum of Quadrant II & III & IV \( (r = .847^*). \) The main reason is that the facilitator did not control the reflection time in Cooperative Puzzle equally among stages and among groups. Moreover, Group 1 and Group 2 had shorter reflection time (approximately 6 minutes), and Group 5 and Group 6 had rather longer reflection time (approximately 13 minutes). The inconsistence of the reflection time had influenced the accuracy of the measure. Therefore, the following discussion will exclude the discussion of Cooperative Puzzle.

| Table 4. Correlation coefficient of Calculator’s the finishing time and the reflection content in storming stage |
|---------------------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|                                          | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  |
| 1. Calculator’s the finishing time        | 1   |     |     |     |     |     |     |     |     |     |
| 2. Quadrant I                             | -.287 | 1   |     |     |     |     |     |     |     |     |
| 3. Quadrant II                            | -.889* | .101 | 1   |     |     |     |     |     |     |     |
| 4. Quadrant III                           | -.083 | .349 | .301 | 1   |     |     |     |     |     |     |
| 5. Quadrant IV                            | -.619 | .153 | -.270 | -.673 | 1   |     |     |     |     |     |
| 6. Quadrant II & III                      | -.699 | .244 | .887 | -.707 | -.125 | 1   |     |     |     |     |
| 7. Quadrant II & IV                       | -.884** | .165 | .663 | -.387 | .900* | .305 | 1   |     |     |     |
| 8. Quadrant III & IV                      | -.817** | .359 | .483 | -.369 | .936** | .180 | .947* | 1   |     |     |
| 9. Quadrant II & III & IV                 | -.980** | .289 | .911 | -.099 | .757 | .554 | .956** | .904* | 1   |     |
| 10. All Quadrant                          | -.976** | .378 | .794 | -.062 | .747 | .559 | .941** | .910** | .996** | 1   |

Note. \( * p < .05; \) \( ** p < .01. \)

The research shows high negative correlation between the finishing times of Calculator in the adjusting stage with Chessboard Maze Quadrant II \( (r = -.889^*), \) Quadrant II & IV \( (r = -.884^*), \) and Quadrant III & IV \( (r = -.817^*). \) In this stage, the facilitator had effectively guided the discussion from Quadrant II to Quadrant IV, and then to Quadrant III; thus, the discussion has reached group development goals. Further, the highest correlation happened in Quadrant II & III & IV \( (r = .980^{**}) \) which shows the game had successfully strengthen group coordination that led to effective group reflections, and resulted in less time to complete the game in the adjusting stage. Due to the frequencies of Quadrant I is close to zero, the results of the Chessboard Maze All Quadrants \( (r = -.976^{**}) \) was ruled out.

The correlation of the total sum of interaction frequency in the storming stage did not reach statistical significance due to the reason that the two games have reversed results that balanced out the outcomes. However, in the result of Chessboard Maze, team members can spend less time to complete the game mission in the adjusting stage if their oral interactions focus on Quadrant II, Quadrant II to III, Quadrant III to IV and Quadrant II to IV and the reflection is conducted in same time by the facilitator in the storming stage.

Results and conclusion

Digital games have the advantages of working out of the limitation of environment and uncontrollable factors such as weather for adventure education courses. Meanwhile, digital games are also a natural attraction to teenagers. Therefore, this research chose six adventure education activities that had counseling theoretical underpinnings and transformed them into digital games. The digital games were cohered into an adventure education course based on Tuckman’s team development model which purpose is to increase group dynamics to enhance group development. Since the main purpose of this research is to understand the effects of digital game-
based adventure education course to groups of various styles, experimental design was not appropriate. Instead, qualitative analysis of the group interactions was conducted with HIM without considering individual factors. All courses were guided by the same facilitator that reflection facilitation style was controlled in the minimal variation.

For research tool – HIM, through HIM and focus group interviews, the reflection sessions after each of the games in the course had significant positive influence to group development, and could improve group performances in the following games. The guiding style of the facilitator affected group development with positive interactions in each of the stages, especially when members faced conflicts in the storming stage (Gilley et al., 2010; Pelegrini Morita & Marie Burns, 2014; Tuckman & Jensen, 1977). From results of HIM, the reflection style focused on the issues of Group (II), Personal (III), and Relationship (IV) was useful for group dynamic and group development (Hill, 1977). For group dynamic, all the six groups finished the first four stages in time. After great reflections, they could solve game missions with better communications and cooperation in next game (Williams & Sternberg, 1988). However, Group 5 and 6 spent half an hour more in the last game to reach the same level of group development. The result was caused by their insufficient interactions in Chessboard Maze at the storming stage. Group 1 to 4 had great peak at the storming stage reversely. All in all, six groups have become effective groups after the course (Johnson et al., 2002).

Though this paper did not investigated the comparison of digital games and traditional adventure education activities, we could refer to the study of Hsu and Shih (2012) to assume that digital games were better than traditional activities among problem-solving and team work. Hsu and Shih (2012) also agreed with the features of digital games were attractive for students. In the past, the studies of team-building focused on human resource and interaction of physics or digital context. However, the course combining adventure education and digital game-based learning in this study was aimed to investigate group dynamic and group development. We expected to provide a digital tool to be applied in adventure education. Nevertheless, this paper confirmed that Tuckman model was also useful in digital game-based learning (Gilley et al., 2010; Pelegrini Morita & Marie Burns, 2014; Tuckman & Jensen, 1977).

The research has drawn to a few conclusions. First of all, using technological tools to add onto the adventure education is feasible and can be successful if designed and guided with proper manner (Hsu & Shih, 2012). Digital learning system can provide simulated environment for the adventure education, but also offer the facilitator important process records for participant observations. In this study, there is high requirement to course assistance. In the future, with system enhancement, working personnel who were responsible for observation and documentation can be reduced. Second, the course was designed with Tuckman’s team development model that the digital game-based adventure education course can appropriately advance group development to become effective groups. Third, the course designed in this research is targeted to team development and human interactions that more learning issues can be explored in the future to test its effectiveness on different variables. Fourth, the games designed in this research can allow participants’ full immersion but some factors were uncontrollable such as unstable network connection and game system bugs that caused occasional game interruptions. Technological improvements can be resolved in the time being. Next, the guiding style of the facilitator would greatly influence the effectiveness of the course and the success of team development. Training to the facilitator to focus on learning issues can be conducted for the courses so that course qualities can be guaranteed. There would be more facilitators available for digital-game based adventure course if training courses can be given. Course structure, course design, and time controls regarding game and the reflection sessions, etc., are to be attended to maintain the quality of treatment. Through the results, it proved that reflection sessions were greatly helpful for group development and positive interaction. Last, the participants chosen in this research are teenagers. In the future, heterogeneity of group members can be tried; the course can also be transformed to adopt children, people with disabilities, or even elders. The games developed in this study are easy to manipulate that can be used in wider range of issues for a variety of participants. The potential of the digital game-based adventure education courses can be extended in the future.

References


Social Networking Sites as Formal Learning Environments in Business Education

Abida Ellahi
Fatima Jinnah Women University, Rawalpindi, Pakistan // abia.ell@gmail.com

(Submitted November 11, 2016; Revised March 14, 2017; Accepted March 16, 2017)

ABSTRACT
The central objective of this study is to investigate the extent to which social networking sites can affect learning effectiveness, and to what amount this technology can be used as supplementary elements for existing pedagogy methods prevailing in a developing country. The study used a teaching case research method to investigate the effects of Social Networking Sites (SNS) usage on learning outcomes of students in higher education setting in Pakistan. The central hypothesis developed in this study was that, using social networking sites in higher education enhance students’ interest, which ultimately increases their satisfaction and perceived learning performance. These effects are further boosted in the presence of instructor’s support. The results confirmed that Social Networking Sites (SNS) hold a place in teaching and learning in higher education. The study provides a way to maximize the impact of the existing technologies, by providing an understanding of how the different technological tools and learning can impeccably be incorporated in higher education.

Keywords
Social Networking Sites, Business education, Satisfaction, Performance, Instructor’s support

Introduction
Instructive technology is in vogue and the recent types of innovation in education have given various chances to scholastic and professional foundations for adopting new forms of learning beyond conventional instruction. The elements found for the most effective learning environments are problem-based learning, collaborative learning, demonstration and application of new knowledge, integration of new knowledge into the learners’ world and assessment of the learning process as well as learners’ progress etc. (Rogers, 2007). The adoption of technologies by academia across the world has reshaped the existing paradigm by bringing the potential solutions for all learning problems. Educational technology has become increasingly common in academia, and governments are spending billions to give academia access to the technology. However, research on the usefulness of using educational technology has lagged in higher education sector. These technologies are not just limited to the use of internet for online literature searching; instead it has moved to the new horizon of web 2.0, digital games, digital badges, learning management systems etc. The technology improves the efficiency and the effectiveness of academic institutional processes in the form of cost and time savings and increases the interaction between institutions and their learners.

The chronicled background of using informational advancement in Pakistan is very recent. Regardless of a few government endeavours, the use of instructive innovation in the nation is still in early stages and it has not achieved its maximum capacity yet. For this country’s case, where there are very few learning technologies existing, normative studies are required to contribute to the understanding and importance of technology management in educational institutes. Since normative studies contribute by bringing new situations that had not been previously investigated. The main problem addressed by this study is the lack of technology usage in higher education of Pakistan. This problem exists because teachers in higher education do not have enough access or exposure to innovative technologies to help them in truly constructing innovative teaching methods. It is, likewise, the need of great importance to recognize what sort of innovation interceded environment works best for learners and what steps ought to be taken to actualize the change. Similarly, it is also necessary to know that how can curriculum areas make use of emerging as well as established technologies to improve students’ interest, engagement, retention and achievement. Hence, careful thinking and research is needed to find the best way to leverage the emerging digital technologies to boost teaching and learning activity.

This study aims to investigate the extent to which new learning technologies like social networking sites can affect learning effectiveness in Pakistan, and to what extent these technologies can be used as supplementary elements for existing pedagogy methods. For this purpose, social networking site has been tested in business education studies. The study arrangement was intended to evaluate the learning satisfaction and perceived learning performance of business graduates with moderating role of instructor’s support because of the utilization of instructive innovations like social networking sites for learning.
Literature review

The efficacious usage of information and communication technologies (ICTs) is not only reliant on the properties of the technology but it must fit in with the social environment in which it is planned to be used (Matzat, 2009). Social networking Sites (SNS) have been used by people as the mainstream communication method to connect with the world including family and friends. It can be considered as one of the paramount way to stay in touch with the members of a society. Academia across the world has not adopted the internet technologies at the same rate. Where some educational institutions have been slow, others have warmly welcomed such technological advancement. The emerging and growing form of internet technologies such as social networking websites or web 2.0, have paved a way to build virtual communities that offer teachers, students and their parents to make real-time interactions.

Social networking websites have been widely accepted by organizations, public, politicians, and media companies. Chu and Du (2013) found that various libraries in academia, adopted SNS to support their work such as for publicity, interacting with library members, and enhancing reference services. Caldwell (2015) claimed that there has been very little research on the usage or potential of using SNS in teaching and learning. Many previous researchers found that SNS had been used by the students only for social purposes. For example, in his study, Stern and Taylor (2007) and Zhao, Grasmuck, and Martin (2008) found that social networking sites have been used by the students to make new friends, send messages and view pictures etc., and later researchers for example, Sanchez, Cortijo, and Javed, (2014) found that students also use social networking sites to share study notes and discuss about the course material and work. In a study conducted by Barczyk and Duncan (2013), students confirmed that “Facebook enhanced their experience of participation in their course” (p. 6).

Constructivism Theory and Social Networking Sites

Constructivism theory argues about providing a learning context in which the learner feels himself in a situation like the one in which he is going to apply the knowledge (Ahdel & Andresen, 2001). It emphasizes on understanding rather than memorizing facts. Lave (1993) argues that in this way, learner constructs, interprets and acts upon the facts being learnt, thus turns out into an enhanced understanding. Socio-culturism approach of learning (Brown, Collins & Duguid, 1989) also argues that learning is enculturation, the process by which the learners become collaborative meaning makers among a group defined by common practices, language, use of tools, values and beliefs. It emphasizes that context around the learning situation is important for effective learning. These theories also support the use of social networking websites (social learning).

Social software support the social constructivist approach to networked learning by providing learners with personal tools that can engage them in social networking (Alexander, 2008). These networking tools enable learners to interact and collaborate in a virtual community during learning activities (Munguautosha, Muyinda, & Lubega, 2011). Due to the personalized and independent learning with respect to place and time, learners are actively involved in the learning activities (Kruger, 2010), thus, it enhances social motivation and interest for students to read and contribute through online discussions. (Walsh, 2010). McLoughlin and Lee (2010) discussed social networking software as a pedagogical choice. They commented that social software facilitates active participation, learner self-direction, and personal meaning construction which is core theme of social constructivism, thus, social software can be applied for teaching.

Theoretical relationships among variables

Learning outcomes

Measuring or determining learning outcomes or effectiveness is a critical part of any learning process. Moody and Sindre (2003) comprehensively defined the two approaches of learning effectiveness which are performance based and perception based assessments of learning. Learning performance is commonly associated with a more positive attitude toward the environment, namely, courses and teachers (Duke, 2002). In this study, students’ perception of satisfaction and perceived learning performance have been taken as dependent variables.
Learning satisfaction and social networking websites

Learning satisfaction in the context of technology infused learning, is the individual’s feelings and experience of learning environment after instruction; hence, it is individual’s joyful feelings or positive attitudes (Lee, 2009). The increasing competitive environment for higher education across the world has brought up students’ satisfaction a key indicator of academic institutes’ effectiveness. Just as low student satisfaction has implications, high student satisfaction can impact a host of outcomes, particularly student retention, student academic success, and social connectedness (Powless, 2011) and institutions’ quality assurance. Al-Rahmi and Othman (2013) investigated students’ satisfaction using social media in higher education and found that in social media interactivity with the teachers, engagement, perceived ease of use, perceived usefulness, interactivity with peers all affect to shape students’ satisfaction in higher education. Hence, it can be hypothesized that H1: Use of social networking sites for learning significantly affects learning satisfaction of students.

Perceived learning performance and social networking websites

According to DeLone and McLean (2002), a successful system impacts individuals by bringing an improvement in their productivity, a change in their operations, as well as a change in understanding the importance and usefulness of the organization’s system. This concept is also captured as perceived individual impact variable which refers to the impact of an information system usage on individual performance of person within his work environment. Helou and Rahim (2014) made an attempt to investigate students’ perceptions about the influence of social networking sites on their academic performance in a Malaysian university. The respondents of their study agreed upon the positive influence of social networking sites on their academic performance. Mehmood and Taswir (2013) explored academic effects of social networking sites on undergraduate students in Oman. They found a significant effect of social networking sites on academic performance of students in Oman. Hence, it can be hypothesized that H2: Use of Social networking sites for learning significantly affects perceived learning performance of students.

Learner’s interest and social networking sites

Effective learning always requires students’ engagement in the learning process as the degree of engagement increase the individual learning outcomes improve as well (Admiraal, Huizenga, Akkerman, & Dam, 2011). Though “individual interest and intrinsic motivation are two separate constructs, they predict similar outcomes in that they both create and sustain a deepening involvement with content and affect over time” (Park, 2012, p. 105). Imlawi, Gregg, and Karimi (2015) argued that although engagement is a significant factor of social networking acceptance, fewer studies have been carried out on the impact of social networking on student’s engagement and perceived educational outcomes. In their study, they found that “instructors who create course-based online social networks to communicate with their students can increase their engagement, motivation, and satisfaction ” (p. 84). Junco, Heibeiger and Loken (2011) also observed that despite an increasing use of social media in education, there is very little research focusing on the effect of social media on students’ interest in higher education. In their controlled experimental study, they used Twitter as tool of social media and found that Twitter usage in an academic way for learning can increase student engagement and their learning performance. They further found that not only students but teachers also become highly engaged when they interact via social media. Hence, social media is an effective educational tool to help student bring the desired results. Hence, it can be hypothesized that:

H 3 (a): Learner’s interest mediates the effect of use of social networking sites on learning satisfaction of students.

H 3 (b): Learner’s interest mediates the effect of use of social networking sites on perceived learning performance of students.

Moderating variable: Instructor’s support in social networking websites

To effectively utilize social networking sites in higher education, Smith (2007) suggested teachers to use popular social networking resources as an extension of the class. He further argued that “by gaining access to social networks in which students are comfortable and already established, connections with those students can be cultivated and developed to facilitate the engagement of students.” (p. 13). Research illustrates an association between academic performance and informal communications between a student and teacher (Terenzini, Pascarella & Blimling, 1999). Mazer, Murphy and Simonds (2007) also indicated that teacher openness via
social networking sites can improve classroom environment and can increase students’ motivation; thus, affecting their learning outcomes. Mirabolghasemi and Iahad (2013) observed that the great potential of social networking sites for learning have been ignored and little attention has been paid to educational benefits of social networking sites. Thus, it has been hypothesized that:

**H 4 (a):** Perceived instructor’s support moderates the relationship between use of social networking websites and learning satisfaction of students.

**H 4 (b):** Perceived instructor support moderates the relationship between use of social networking websites and learning performance of students.

The proposed research model is shown in Figure 1.

![Proposed research model](image)

**Figure 1.** Proposed research model

**Research method**

Methodologically, this study falls within the broad framework of case study method. Aaltio and Heilmann (2010), a case study is a “special research strategy and approach that can use either qualitative or quantitative data, or even combinations of them” (p. 65). This study represents a teaching methodology case in a university setting. Brem (2010) highlighted the importance of teaching cases as “Nevertheless, teaching cases offer a highly effective method for including students in the learning process and making a transfer between theory and practice. More research is needed regarding the success factors of teaching outlines, moderators, and group sizes” (p. 77). This study used a quantitative single case research in a university classroom. Boyer (2010) in an encyclopedia of case study research described that “quantitative single-case research is an experimental design that can be conducted with one subject or an entire group treated as a subject. The quantitative single-case research design (QSCRD) is relevant to case study research because it is a strategy used to identify a causal relationship between variables for one subject or individual subjects” (p. 757).

**Study setting**

There has been a growing emphasis on educational technologies for learning and teaching. Although the learning technologies like Web 2.0 and digital games etc. are promising in nature, most of the universities and colleges have not yet adopted these technologies in Pakistan. For a learning activity in University, where assignments are submitted online or typed on a computer, and soft copies of online notes and slides are prepared, a consolidated platform is required where all the learning material can be kept in one place for easy access. As a teacher, researcher of this study felt that her University lacked a learning management system. Teacher and students usually used yahoo or google groups for sharing assignments and notes. Although these groups facilitated the interaction but they also have numerous limitations. Secondly, it was also a common observation that students or young generation now spend more time on social networking websites due to their interactive features. It was also observed that due to large number of class students in a single course, every student does not have an equal access to participate in the class or to communicate with the teacher. Most of the communication was held via a focal student who is a class representative and has a regular contact with the teacher. Communication outside the class through only a single person created many problems. Thus, affecting students’ learning satisfaction and performance. At the University level, there was no platform or communication medium that could equally involve all the students in communicating with teacher at the same time. The lack of a technological system posed a need to establish or set one that can aid in both teaching and learning process in a course. After
searching, it was easy to set up a social networking site that could connect a class group not only during class but also beyond the class walls.

**Edmodo: The SNS platform**

It was planned to adopt a social network site to address the above-mentioned problems. After careful evaluation, Edmodo was chosen to be implemented in the course. This author was the course instructor and this was the first time when students in this department were given a supportive resource to enhance their learning. The study was conducted from February to June, 2015, from September to December, 2014 and from February to June 2015 in three consecutive semesters at a public-sector university. The participants in this study were students enrolled in an undergraduate and graduate level courses of “IT in Business and Computing Skills for Management” (as shown in Figure 2). The syllabus of these courses was oriented towards technology usage in business which meant that most of the activities of these two courses were web and computer based. The classes of these two courses were held in a lab room of the department, where students had access to computers and internet.

![Screen shots of Edmodo](image)

**Figure 2. Screen shots of Edmodo**

**Procedure**

At the beginning of each semester (from February 2014 to June 2015), the teacher integrated one course with Edmodo where students enrolled in that course were asked to register and join the online classes. In this way, the courses were blended with one to one teaching and with the online collaboration throughout each semester. To benefit from Edmodo, space for learner’s centered exploration and discussion was created. Students were informed about the usage of this platform and were briefed about the assignments, quizzes and announcements made via this platform. Students were encouraged to share study related information to the group where other students could see, like and comment on it. Thus, a full collaborative environment was established where each student could see the other’s progress anywhere by logging in to this network.

**Data collection**

After successfully running the SNS platform of Edmodo for three consecutive semesters, questionnaires (surveys) as a tool of data collection were distributed to the students to gather their perceptions and experiences of using the Edmodo platform. The questionnaire was basically employed to investigate the students’ opinion towards Edmodo as well as to get their suggestions. This study adopted the measures used to operationalize the constructs included in the previous literature, making minor wording changes to tailor these measures to the current context. The questionnaires were refined through pre-testing. The pre-testing was focused on instrument clarity, question wording and validity. During the pre-testing, 10 users of the social networking sites were taken
as subjects and were invited to comment on the questions and wordings. The comments of these 10 subjects were taken as a basis for the revisions of construct measures.

**Sampling**

The general population for this study were students who use social networking websites. The sample from this population was drawn based on purpose or judgment i.e., to include those students who used Edmodo website as SNS platform for learning and education.

**Table 1. Mean, standard deviation, alpha and correlation**

<table>
<thead>
<tr>
<th>Variable</th>
<th>M</th>
<th>SD</th>
<th>α</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNS Use</td>
<td>2.363</td>
<td>.791</td>
<td>.92</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learner Interest</td>
<td>2.535</td>
<td>.918</td>
<td>.92</td>
<td>.718*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning Satisfaction</td>
<td>2.365</td>
<td>.894</td>
<td>.94</td>
<td>.825* .775**</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Learning Performance</td>
<td>2.453</td>
<td>.825</td>
<td>.90</td>
<td>.834** .775** .880**</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Instructor’s support</td>
<td>2.396</td>
<td>.802</td>
<td>.88</td>
<td>.577** .555** .578** .636**</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Note. Pearson Correlation, Sig. (2-tailed), N = 127; *Correlation is significant at the 0.05 level (2-tailed), **Correlation is significant at the 0.01 level (2-tailed).

In this study, purposively or deliberately only those students were included for survey, who attended a course integrated with Edmodo platform. Total 150 students who were enrolled in those three courses in three semesters were selected. Questionnaires were self-administered to them which were filled by them anonymously. Out of 150, 127 useful responses were received that were analyzed for this study.

**Results**

Data was collected using a questionnaire based on previous research. The collected responses were analysed with the help of statistical tests using SPSS software.

**Descriptive statistics**

The statistical values in Table 1 give information about means, standard deviations, alpha and correlations among the variables. The means and standard deviation values for SNS use were \( M = 2.3, SD = 0.79 \), for learner’s interest \( M = 2.5, SD = 0.918 \), for learning satisfaction \( M = 2.3, SD = 0.89 \), for learning performance \( M = 2.4, SD = 0.825 \) and for instructor’s support were \( M = 2.3, SD = 0.802 \). Correlation is used to show the relationship between to variable. A Pearson correlation coefficient was computed to assess the relationships among the variables. The highest correlation was observed between variables learning performance and learning satisfaction \( r = 0.880, p = .000 \) and the lowest correlation was observed between variables learner’s interest and instructor’s support \( r = 0.555, p = .000 \). All correlations above 0.10 were significant at \( p < .01 \) level. All variables were significantly and positively related with other. The Cronbach alpha values, a measure of scale reliability are also given in Table 1. The alpha values for ten items variable SNS use were 0.82, for eight items variable learner interest were 0.92, for eight items variable learning satisfaction 0.94, for eight items variable learning performance 0.90 and for six items variable instructor’s support were 0.88. All the values show excellent level of scale reliability.

**Hypothesis testing**

The Hypotheses were tested with the help of SPSS process Macro by Hayes and Preacher (2014). Hypotheses were constructed on the bases of theoretical background. To test the acceptance and rejection of the mediation hypotheses, regression using PROCESS macro model 4 was applied.

**Mediation analysis**

The statistical values shown in table 2 present the direct and indirect effect of independent variables on dependent variables. The values indicated that SNS use is a significant predictor of learning satisfaction and it
accounts for 93 units change in learning satisfaction ($b = 0.93$, $t (125) = 16.33$, $p = .000$). The results also confirmed that the effect of SNS use on perceived learning performance, ignoring the learner’s interest was significant ($b = 0.86$, $t (125) = 16.87$, $p = .000$). In both cases, the SNS use was also a significant predictor of learner’s interest which brought 83 units change in learner’s interest ($b = .83$, $t(125) = 11.53$, $p = .000$).

The results of indirect effect or mediation analysis indicated the learner’s interest as a mediator that was significantly related to learning satisfaction ($b = 0.48$, $t(124) = 8.58$, $p = .000$). It was observed that the effect of SNS use was reduced from $b = 0.93$ to $b = 0.53$ with the inclusion of learner’s interest ($b = 0.53$, $t(124) = 8.13$, $p = .000$). Hence, partial mediation can be observed in this analysis.

In the context of learning performance, mediation analysis shows that the mediator (learner’s interest) has significant effect on dependent variable (perceived learning performance). It showed that with the inclusion of the mediator, effect of independent variable, SNS use, on perceived learning performance was still significant. However, the intensity of the effect was lessened from $b = 0.86$ to $b = 0.59$. Hence, partial mediation can be observed in this analysis.

The results of Sobel test (Normal theory tests for indirect effect) also found mediation in the model for learning satisfaction ($z = 6.86$, $p = .000$) and ($z = 5.14$, $p = .000$) for learning performance. The indirect effects were also significant at the 95% level of significance, as indicated by the values of LLCI and ULCI when the lower and upper levels of the confidence intervals did not show zero. The partial mediation in both cases implies that there is not only a significant relationship between learner’s interest and learning satisfaction as well as between learner’s interest and learning performance, but also some direct relationship between SNS use and learning satisfaction as well as between SNS use and learning performance. Hence, Hypotheses H1, H 2, H 3 (a) and H 3 (b) have been accepted.

### Table 2. Mediation analysis

<table>
<thead>
<tr>
<th>Path</th>
<th>Predictors</th>
<th>$\beta$</th>
<th>SE</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Path $a$</td>
<td>SNS Use --- Learner’s Interest</td>
<td>.83</td>
<td>.072</td>
<td>11.53</td>
<td>.000</td>
</tr>
<tr>
<td>2 Path $b$</td>
<td>Learner’s Interest --- Learning Satisfaction</td>
<td>.48</td>
<td>.056</td>
<td>8.58</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>Learner’s Interest --- Learning Performance</td>
<td>.32</td>
<td>.056</td>
<td>5.76</td>
<td>.000</td>
</tr>
<tr>
<td>3 Path $c$</td>
<td>SNS Use ---Learning Satisfaction</td>
<td>.53</td>
<td>.065</td>
<td>8.13</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>SNS Use ---Learning Performance</td>
<td>.86</td>
<td>.051</td>
<td>16.87</td>
<td>.000</td>
</tr>
<tr>
<td>4 Path $C^\prime$</td>
<td>SNS Use---Learning Satisfaction</td>
<td>.93</td>
<td>.057</td>
<td>16.33</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>SNS Use---Learning Performance</td>
<td>.59</td>
<td>.065</td>
<td>9.03</td>
<td>.000</td>
</tr>
<tr>
<td>Bootstrap Results for Indirect Effect (Learning Satisfaction)</td>
<td>Effect</td>
<td>SE</td>
<td>LLCI (95%)</td>
<td>ULCI (95%)</td>
<td></td>
</tr>
<tr>
<td>Learner’s Interest</td>
<td>.3357</td>
<td>.0605</td>
<td>.2436</td>
<td>.4788</td>
<td></td>
</tr>
<tr>
<td>Bootstrap Results for Indirect Effect (Learning Performance)</td>
<td>Effect</td>
<td>SE</td>
<td>LLCI (95%)</td>
<td>ULCI (95%)</td>
<td></td>
</tr>
<tr>
<td>Learner’s Interest</td>
<td>.2619</td>
<td>.0569</td>
<td>.1465</td>
<td>.3635</td>
<td></td>
</tr>
</tbody>
</table>

Note. Dependent Variable: Learning Satisfaction and Learning Performance, LL = lower limit; UL = upper limit. N = 127; Unstandardized regression coefficients are reported.

### Moderation analysis

To test the moderation effect of instructor’s support on learning satisfaction and perceived learning performance, model one of Process macro developed by Hayes and Preacher, (2014) was used. Table 3 shows that the effects of instructor’s support ($β = 0.1623$, $t(123) = 2.37$, $p = .01$) and SNS use ($β = 0.81$, $t(123) = 11.40$, $p = .000$) on learning satisfaction were significant. However, when interaction term was regressed, it did not produce significant effect ($β = 0.04$, $t(123) = 0.78$, $p = .43$) which can be regarded as no moderation effect was proved. Table 3 also shows that the effects of instructor’s support ($β = 0.23$, $t(123) = 3.9$, $p = .0002$) and SNS use ($β = 0.71$, $t(123) = 11.46$, $p = .000$) on perceived learning performance were significant. However, when interaction term was regressed, it again did not bring out significant effect ($β = 0.051$, $t(123) = 0.95$, $p = 0.34$) which can be regarded as no moderation effect was proved. As the moderation analysis did not prove the hypothesis, hence
further examination of the interaction effect by plotting the slopes and high and low levels (1 SD above and below the mean) was not conducted. Hence hypotheses H 4(a) and H 4(b) have been rejected.

Table 3. Moderation analysis of learning satisfaction

<table>
<thead>
<tr>
<th></th>
<th>Model summary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R</td>
</tr>
<tr>
<td>Model 1 -- Learning Satisfaction</td>
<td>.8354</td>
</tr>
<tr>
<td>Model 2 -- Learning Performance</td>
<td>.8561</td>
</tr>
</tbody>
</table>

Path coefficients

<table>
<thead>
<tr>
<th></th>
<th>β</th>
<th>SE</th>
<th>t</th>
<th>p</th>
<th>LLCI</th>
<th>ULCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Instructor’s support -- Learning Satisfaction</td>
<td>.1623</td>
<td>.0683</td>
<td>2.3754</td>
<td>.0019</td>
<td>.2975</td>
<td>.0270</td>
</tr>
<tr>
<td>Instructor’s support -- Learning Performance</td>
<td>.2311</td>
<td>.0592</td>
<td>3.9017</td>
<td>.0002</td>
<td>.3483</td>
<td>.1138</td>
</tr>
<tr>
<td>2 SNS use -- Learning Satisfaction</td>
<td>.8170</td>
<td>.0716</td>
<td>11.4097</td>
<td>.0000</td>
<td>.9587</td>
<td>.6752</td>
</tr>
<tr>
<td>SNS use -- Learning Performance</td>
<td>.7119</td>
<td>.0621</td>
<td>11.4680</td>
<td>.0000</td>
<td>.8348</td>
<td>.5890</td>
</tr>
<tr>
<td>3 SNS use * Instructor’s Support (Learning Satisfaction)</td>
<td>.0486</td>
<td>.0616</td>
<td>.7890</td>
<td>.4317</td>
<td>.1704</td>
<td>.0733</td>
</tr>
<tr>
<td>SNS use * Instructor’s Support (Learning Performance)</td>
<td>.0510</td>
<td>.0534</td>
<td>.9550</td>
<td>.3415</td>
<td>.1566</td>
<td>-.0547</td>
</tr>
</tbody>
</table>

R²-change due SNS use * Instructor’s Support (Learning Satisfaction) = 0.0015
R²-change due to SNS use * Instructor’s Support (Learning Performance) = 0.0020

Discussion of findings

The study aimed to examine the effects of social networking websites platform on learning satisfaction and learning performance in higher educational settings. The use of social networking sites over a period of three consecutive semesters provided some useful insights from the study which are summarized in this section.

The study was a preliminary research which investigated the effects of SNS use on learning satisfaction and perceived learning performance of students in higher education with a mediating role of learner’s interest and moderating role of instructor’s support. The statistical results confirmed that SNS use in this study positively affected learning satisfaction and perceived performance of students.

The findings of this study are consistent with many previous studies. For example, in a study AL-Rahmi and Othman (2013) investigated students’ satisfaction using social media in higher education and found that in social media interactivity with the teachers, engagement, perceived ease of use, perceived usefulness, and interactivity with peers all affect to shape students’ satisfaction in higher education. They also found that collaborative learning in social media enhances student academic performance and SNS use creates interest among learners and their interest in learning shapes their learning satisfaction and performance. Thus, confirming the mediating role of interest in this theoretical model. The study found a partial mediation of learner’s interest which implies that there is not only a significant relationship between the learner’s interest and the learning satisfaction and performance, but also some direct relationship between the SNS use and learning satisfaction and performance. The results are again consistent with previous studies where Mbodila, Ndebele and Muhandji (2014) also confirmed the influence of using social networking websites on students’ learning and engagement (interest) in higher education.

When instructor’s support was regressed on learning satisfaction and perceived learning performance, it brought a positive effect on learning satisfaction and perceived learning performance. However, when interaction term was added to check the moderating effect, it did not bring out significant result. These findings are not consistent with many previous studies like den Exter, Rowe, Boyd and Lloyd (2012) who found the importance of role of teacher guidance for the success of this emergent approach of using Web 2.0 technology and of Ke (2010) who strongly advocated the presence of teachers for adult students in web learning, as adult students show more learning satisfaction in an instructor’s continuous presence. Ford and Lott (2009) said that in constructivist learning environment, students are responsible for learning and teacher’s role is of a facilitator. In their words “the social interactions between students and students, and teachers and students have changed. Students are no longer dependent on the teacher as the main source of information.”

These findings do not show that instructor’s support was a useless variable, in fact, its moderating effect was not proven. The non-significance of this moderator variable does not mean that there is zero effect of instructor’s support in Web 2.0 learning environment. It might indicate that the study lacked the statistical power to
distinguish the value from zero. A possible explanation of this insignificant results is that moderation was checked on direct effect of SNS use on learning satisfaction and performance. However, this could also be checked in the form mediated moderation and moderated mediation. Defining the path in this way might provide the instructor’s support as a significant moderator in the study. The finding is that instructor’s support is a valid and strong predictor of learning satisfaction and performance. However, path defining its moderating role needs to be re-devised.

The use of social networking sites provided the learners an opportunity to construct their knowledge and engage themselves in this learning process. In the context of social networking sites, students could construct their own knowledge, share and collaborate with their peers. They could see the progress of other students; thus, social aspect of learning was included in this technology. In this way, learning was an active process and student centred approach was evident in this case. Teachers’ role was as a facilitator in this case. In consistency with theory and practice, students were expected to expand their knowledge beyond the walls of the class. Hence, in the light of constructivism theory, they could learn in an active and social process. They were engaged in “active dialogue” with their fellows and the teacher. Hence, an interactive learning community was established that facilitated better learning outcomes. The learning outcomes in the form of students’ satisfaction and perceived learning performance from students’ opinions is evident that social networking sites promote and support constructive learning which strengthens their learning outcomes. Overall, the students found the use of social networking tool as helpful in their learning and it had a positive effect on their learning satisfaction and performance.

Limitations and future recommendations

Though this study could unveil some interesting and meaningful results, there are still some limitations and like any other study, this study is not an exception. The first limitation of this study was that it was conducted in a single university and context. The implementation of such a study across many departments and universities could expose more angles of this phenomena. The second limitation of this study was inclusion of students in an IT course where students could take it mandatory as part of learning technology for a course. Hence, involvement of students from other courses of social sciences would shed more light on this problem area. Another limitation was inclusion of only one gender, i.e., female which could have caused a gender bias. The investigation of gender role in using IT has been evident in previous studies. However, this part was beyond the control, as the research method adopted in this study required no inclusion of gender effect. This study of social networking sites provides a baseline to further investigate the potential of Web 2.0 technologies in academia. The study triggers the need for more investigation on the topic by investigating further research constructs to better predict technology integration process in academia and exploit them as active teaching and learning tools. Considering this study’s findings, the same method and research design can be applied to different contexts across other universities and study disciplines with larger sample size and diverse learner groups. The significant and non-significant variables in this study can be further examined in other studies. Future research can be conducted to find the role of individual difference in gender, personality, type of academic in using social networking sites for learning.

Implications

Undoubtedly, as evident in this research, there are so many technological tools available that can boost learning and teaching in higher education. However, the remote control of using such technologies is in the hands of teachers whose willingness and adoption of technologies determine the actual success of technology infused learning. The findings from this study have important implications for teachers, higher education policy makers, trainers as well as software designers. By taking Edmodo as a successful learning case in this study, teachers can do more to utilize other productive web 2.0 tools to enhance collaborative learning among their students that goes beyond the classroom. Teachers can also innovate their teaching styles offering their students more novel opportunities to learn. However, teachers need to define their role and level of involvement in online community while blending the classroom with web 2.0. Teachers or practitioners can use social media in classroom to blur conventional communication among teachers and students by promoting their passion for learning. The findings of this study confirmed an increase in students’ learning outcomes, hence, teachers can use web 2.0 technologies to help shy students feel more confident, promote learning among peers and enhancing students’ self-efficacy.

Software designers could introduce specialized social networking software for learning in each discipline. They can also integrate social networking site features into existing learning management systems. Many important
insights can also be yielded for instructional designers and scholars who are interested in integrating online collaborative communities to support learning. For them, a thorough understanding of teacher’s role could help to find a way to perfectly align the new technology with existing pedagogical methods. For policy makers and instructional designers, it is necessary to make sure the availability of technical resources and support at campus for a productive blended classroom.

Conclusion

The findings of this study confirmed that students in higher education sector in Pakistan are ready to accept and adopt new technologies that can better facilitate their learning process. They are ready to create and share their knowledge in a collaborative manner by using technological platforms. The outcomes of using Web 2.0 tool in education are encouraging for teachers and practitioners who want to experience and facilitate their teaching process with new emerging technologies. It can be concluded that social networking sites hold a place in teaching and learning in higher education. The use of social networking sites provides a holistic learning process, and to endorse their values in education, they must be tested as an opportunity. The conclusion of this study insists on the utility and necessity of constructive and collaborative learning beyond class boundaries using the web 2.0 tools. Students must be allowed and facilitated to share their own generated content to strengthen their learning. To meet the global challenges and development in higher education, educational institutes must adopt these technologies. Hence, it is vital to take advantage of the software which are freely available to remain competitive by providing world class education. The higher education sector in Pakistan must not consider it as simply buying and installing the software in universities; rather, it should be regarded as a social change in higher education sector “an indispensable aim to shape a generation of young learners.”

References


Factors Related to ICT Competencies for Students with Learning Disabilities

Ting-Fang Wu¹, Cheng-Ming Chen², Hui-Shan Lo³, Yao-Ming Yeh⁴ and Ming-Chung Chen⁵*

¹Graduate Institute of Rehabilitation Counseling, National Taiwan Normal University, Taiwan // ²Department of Special Education, National Kaohsiung Normal University, Taiwan // ³Department of Special Education, National Taiwan Normal University, Taiwan // ⁴Department of Information Management, Kainan University, Taiwan // ⁵Department of Special Education, National Chiayi University, Taiwan // tfwu@ntnu.edu.tw // tomsong2@ms18.hinet.net // 80209003e@ntnu.edu.tw // ymyeh@mail.knu.edu.tw // michen@mail.ncyu.edu.tw

*Corresponding author

(Submitted November 16, 2016; Revised January 17, 2017; Accepted April 15, 2017)

ABSTRACT

The purpose of this study was to develop an explanatory model for the information and communication technology (ICT) competencies of students with and without learning disabilities (LD). A conceptual model was proposed, and included five major constructs: (1) attitude towards using ICT, (2) ICT competency, (3) demographic characteristics, (4) ICT access, and (5) purpose for using ICT. A self-reported questionnaire, Scale of Digital Participation (SDP), was designed by the authors and used to collect data. Data from 117 elementary school students with LD and 117 without LD were used to test the initial model. Meanwhile, data from 102 junior high school students with LD and 102 students without LD were used to evaluate the cross-validation of the revised model. The results of structural equation modelling (SEM) demonstrate acceptable goodness of fit for the initial ICT competency model but poor parsimony. Therefore, a revised ICT competency model with both parsimony and goodness of fit was generated using data from elementary school age students. Moreover, this new model was also able to further explain the complex phenomena of ICT competency for junior high school age students. Nevertheless, no differences were found between the elementary school and junior high school models, thereby confirming the cross-validation of the revised model.

Keywords

Information and communication technology, Students with learning disabilities, Structural equation modelling

Introduction

Information and communications technology (ICT) has been regarded as an equalizer for students with disabilities participating in learning activities (de Jong, Specht, & Koper, 2010; Huang et al., 2010; Tosun & Baris, 2011). Students with disabilities are enabled to participate in learning and communication with the support of ICT (Fichten et al., 2009; Ko, Chiang, Lin, & Chen, 2011). However, with the popularity of ICT use in our daily life, many barriers continue to exist, prohibiting some students from participating in learning. Thus, the digital divide has become yet another barrier facing students with disabilities.

The digital divide has been defined as the lack of opportunity to access ICT and the competencies to effectively use ICT (Mäkinen, 2006). Access and ability are two major issues when exploring the digital divide (DiMaggio, Hargittai, Celeste, & Shafer, 2004; Hargittai, 2003; Latimer, 2001; Mossberger, Tolbert, & Stansbury, 2003; Stanley, 2003; van Dijk, 1999; van Dijk & Hacker, 2003; Warschauer, 2003). Access is defined as the actual availability of and access to ICT devices. Ability is a term that covers both an individuals’ ability to cope with ICT devices and the amount of help and support individuals receive from their social network (Lebens, Graff, & Mayer, 2009).

In addition to ICT access and ability, Martin (2003) adds motivation as another important factor necessary for using ICT to learn. Motivation refers to the individual’s attitude towards learning through the use of ICT and the willingness to dedicate one’s effort to acquire ICT competencies.

Disability has been emphasized as one factor that can impact individuals’ access to ICT (Department for Education and Skills, 2001; Russell & Stafford, 2002). Results from the literature have shown that persons with disabilities have fewer opportunities to access ICT (Department for Education and Skills, 2001; Russell & Stafford, 2002; U.S. National Telecommunications and Information Administration, 2011). However, previous surveys on the digital divide have overlooked students with learning disabilities (LD) (U.S. National Telecommunications and Information Administration, 2011; Research, Development and Evaluation Commission, 2012), even though the students with LD represent approximately 5% of school-aged students.
(Hallahan, Lloyd, Kauffman, Weiss, & Martinez, 2004). Only a few studies have focused on this population (Chen et al., 2014; Wu, Chen, Yeh, Wang, & Chang, 2014). Wu and her colleagues (2014) investigated 117 students with LD and their non-LD peers from 4th to 6th grade. The results indicated no significant difference in opportunities to access computers and the Internet, either at home or at school, between children with and without LD. However, the ICT competencies of children with LD were significant poorer than their non-LD peers. Chen and his colleagues’ (2014) investigation also supported the existence of a difference in ICT competencies between students with and without LD.

While it seems to be a fact that a significant difference in ICT competencies exists between students with and without LD, the factors contributing to this phenomenon remain under-researched. Previous research does indicate that age, gender, demography, social economic status (SES), and disabilities are regarded as major factors related to the digital divide (Enoch & Soker, 2006; Lebens et al., 2009; Martin, 2003; Mckenzie, 2007; Stoiles & McDougall, 2011; van Deursen, van Diik, & Peters, 2011; Vicente & López, 2010). Other recent studies have tended to focus on fluid variables, such opportunities for using ICT, or level of education, particularly when focusing on school-age children (Wu et al., 2014).

Meanwhile, the digital divide remains a complex concept, which is comprised of a variety of factors, including attitude, ability, and access. Which one should be the focus? While ability or competence might be the core of this complex concept, studies have also shown that access is one of the key factors influencing ICT competence (e.g., van Deursen et al., 2011). In addition, due to the diffusion of ICT in developed and developing countries, the importance of access has tended to be replaced by that of competency (van Dijk, 2006; Warschauer, 2003).

Since the digital divide is a complex phenomenon, it is impossible to explore it simply through bivariate analysis, as this basic approach not only increases the type I error rate, but is unable to deal with latent variables. Therefore, structural equation modeling (SEM) is deemed a proper solution for testing conceptual or theoretical models which are comprised of many variables. In addition, SEM is able to analyze latent variables and observed variables simultaneously. Therefore, SEM has been widely adopted to analyze complex phenomenon (Bowen & Guo, 2012), including recent digital divide research (e.g., Constantin, Taylor, Park, & Cho, 2006; van Deursen et al., 2011).

Some studies which have adopted available theoretical frameworks, such as the Technology Acceptance Model (TAM) (e.g., Tarhini, Hone, & Liu, 2014) for their research. However, this study, as with van Deursen et al. (2011), proposed an initial model which first included possible factors found in the literature, and then excluded insignificant factors and paths to create a revised SEM model based on best analytical results. However, could the model generated from SEM testing be useful and explanatory when applied to a different data set? Previous studies seem to stop testing the revised model (e.g., Tarhini et al., 2014) or to use half of the original sample to test the revised model (e.g., Igbaria, Guimaraes, & Davis, 1995). However, cross-validation should be further tested to demonstrate the effect of the revised model across different groups (MacCallum, Roznowski, Mar, & Reith, 1994).

Based on the abovementioned considerations, the purpose of the current study was to develop an explanatory SEM model to represent the relationships among key ICT competency factors. The authors first designed an initial conceptual model which included all the proposed factors generated from previous studies. The theoretical framework is shown in Figure 1, and includes five major constructs, (1) attitude towards using ICT (confidence and motivation), (2) ICT competency (basic computer skills, Office software use, and Internet use), (3) demographic characteristics (disabilities, grade), (4) ICT access (family ownership of a computer, frequency of Office software use) and (5) purpose for using ICT (learning, social interaction, leisure, and daily life needs), proposed as a theoretical conceptual model. The significance of paths between factors were tested, and only significant paths were included in creating an increasingly parsimonious revised model. Additionally, in order to explore cross-validation, a sample from another group of participants was used to examine the goodness of fit of the revised SEM model. Furthermore, tests of measurement invariance were conducted to compare the two models. As such, the specific research questions of this study are:

- Could the five proposed factors related to ICT competency be used to develop an SEM model which could explain these complex phenomena and the relationships among factors for elementary students with and without learning disabilities?
- Would the final revised SEM model also explain the complex phenomena of ICT competency factors for junior high school students with and without learning disabilities?
- Would differences between SEM models for elementary school and junior high school students be found?
The specific hypotheses in this study are the following:

Hypothesis 1a: Students without LD will have a better attitude towards using ICT than students with LD.

Hypothesis 1b: Higher grades will be associated with more positive ICT attitude.

Hypothesis 2a: Students without LD will have higher levels of ICT competency than students with LD.

Hypothesis 2b: Higher grades will be associated with higher levels of ICT competency.

Hypothesis 3a: Students who have a computer at home will have more positive ICT attitudes than students without a computer at home.

Hypothesis 3b: The frequency of using Office software will be associated with more positive ICT attitude.

Hypothesis 4a: Students who have a computer at home will have higher levels of ICT competency.

Hypothesis 4b: The frequency of using Office software will be associated with higher levels of ICT competency.

Hypothesis 5a: Higher levels of using ICT for learning activities will be associated with more positive ICT attitudes.

Hypothesis 5b: Higher levels of using ICT for social interaction will be associated with more positive ICT attitudes.

Hypothesis 5c: Higher levels of using ICT for leisure activities will be associated with more positive ICT attitudes.

Hypothesis 5d: Higher levels of using ICT for daily life activities will be associated with more positive ICT attitudes.

Hypothesis 6a: Higher levels of using ICT for learning activities will be associated with higher ICT competency.

Hypothesis 6b: Higher levels of using ICT for social interaction will be associated with higher ICT competency.

Hypothesis 6c: Higher levels of using ICT for leisure activities will be associated with higher ICT competency.

Hypothesis 6d: Higher levels of using ICT for daily life activities will be associated with higher ICT competency.

Hypothesis 7: Students’ with more positive ICT attitude will demonstrate higher levels of ICT competency.
Methods

Research design

The purpose of this study was to develop a model which could explain the relationships among ICT competency and other proposed factors. The authors first developed a theoretical conceptual model (see Figure 1). Then, data from a previous study (Wu et al., 2014) was used to test the initial model and to generate a revised model. Finally, data from another sample was used to test the revised model. Therefore, two studies were conducted. The first study investigated the relationships among the abovementioned five constructs. The resulting analyses were then used to revise the theoretical model to create a revised model. In the second study, data generated from another sample was used to examine the goodness of fit of the revised model. Then tests for measurement invariance were conducted to compare the relative fit of the two models.

Participants

In the first study, data were reanalyzed from a previous study (Wu et al., 2014), adopting paired sampling based on gender and grade. 234 students from elementary schools in Taiwan were recruited, including 117 students with LD and 117 non-LD peers (NLD). To test the cross-validity of the revised model, 204 students from junior high schools in Taiwan (LD = 102, NLD = 102), participated in the second study. The number of participants is listed on Table 1.

Participants with LD were recruited through telephone calls to special education teachers of elementary and junior high schools in Taiwan. They were asked to invite their students with LD to participate in this investigation. Students with LD were diagnosed and identified by the local education authorities based on the specific criteria: (a) normal intelligence quotients; (b) significant intra-individual differences among skills; and (c) significant difficulties in academic performance, such as listening comprehension, oral expression, word recognition, reading comprehension, writing, and calculation, were not improved after interventions provided through regular instruction. Furthermore, no neurological deficits, intellectual delay, physical impairments, or cultural disadvantage were reported. All participants with LD were placed in regular classes with their non-disabilities peers. The non-LD peers were also recruited from the same schools. Parental consent was obtained before the investigation.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Male</th>
<th>Female</th>
<th>Subtotal</th>
</tr>
</thead>
<tbody>
<tr>
<td>4th</td>
<td>23</td>
<td>19</td>
<td>42</td>
</tr>
<tr>
<td>5th</td>
<td>28</td>
<td>12</td>
<td>40</td>
</tr>
<tr>
<td>6th</td>
<td>26</td>
<td>9</td>
<td>35</td>
</tr>
<tr>
<td>7th</td>
<td>15</td>
<td>9</td>
<td>24</td>
</tr>
<tr>
<td>8th</td>
<td>31</td>
<td>6</td>
<td>37</td>
</tr>
<tr>
<td>9th</td>
<td>30</td>
<td>11</td>
<td>41</td>
</tr>
</tbody>
</table>

Table 1. The number of students with and without LD

Instruments

A self-developed questionnaire, Scale of Digital Participation, was adopted. For elementary school students, the questionnaire was titled as Scale of Digital Participation for Elementary School Students, while for junior high school students, the title of the instrument was Scale of Digital Participation for Junior High School Students. The questionnaires contained identical items in five major sections: personal demographic information, current status of ICT access, purpose for using ICT, ICT competency, and ICT attitude.

Personal demographic information included grade and LD or non-LD status. ICT access focused on opportunities to use ICT and ownership of ICT equipment (e.g., frequency of Office software use, and family ownership of a computer). Since ICT access for students with LD did not differ from their non-LD peers (Wu et al., 2014), this paper used family ownership of a computer and frequency of using Office software alone to represent ICT access.

This study considered four major purposes for using ICT: learning, social interaction, leisure, and daily life needs. Using ICT for learning includes using ICT to learn foreign languages and to complete assignments. Using
ICT for social interaction includes using social media to interact with friends, such as through Facebook. Using ICT for leisure includes playing games and listening to music. Finally, using ICT for daily life needs consists of searching for information for train schedules, or online shopping.

The subscale for ICT competency used in the previous study consisted of six types of computer skills, which included basic computer operation (10 items), word processing (10 items), spreadsheet software use (10 items), presentation software use (7 items), graphic software use (4 items), and Internet use (10 items) (Wu et al., 2014). In this study, there were too many types of ICT skills for inclusion in the SEM model. Therefore, we combined word processing (10 items), spreadsheet software use (10 items), and presentation software use (7 items) into a new category named Office software use (27 items). Then, Office software use, basic computer operation (10 items), and Internet use (10 items) were included in the model. Each item comprised a simple ICT task. For example, the item “Can you download a file from a website?” is one of the items used for assessing Internet use skills. It is a yes/no question. If the student indicates the ability to perform this task, one point was scored for this item. The average points, which ranged from 0.00 to 1.00, were used to represent ICT competency for each category, collectively named as the Competency Index (CI). Higher CI represented greater competency (Wu et al., 2014).

The subscale of ICT attitude includes two concepts, motivation (4 items) and confidence (2 items) in using ICT. Motivation towards using ICT included four items. For example, “I am interested in learning knowledge related to computer and internet use” was one of the motivation items for attitude towards ICT.

Wu and her colleagues (2014) established content validity and internal consistency for the questionnaire adopted in this study. The internal consistency of each category of ICT competency and attitude, represented by Cronbach’s α, were reliable (basic computer operation = .80, Office use = .95, Internet use = .89, motivation = .42; confidence = .39). Test-retest reliability was established through the use of a Pearson product-moment correlation coefficient, by completing the questionnaire a second time after a two week interval, with the test-retest reliability results as follows: basic computer operation \( r = .75, p < .01 \); Office use \( r = .80, p < .01 \); Internet use \( r = .83, p < .01 \); motivation \( r = .46, p < .01 \); and confidence \( r = .62, p < .01 \). Thus, the statistical results demonstrate the sufficient reliability of the instrument.

**Procedure**

Printed questionnaires were delivered to the sampled schools. A letter of explanation was also attached. The procedure for completing the questionnaires was explained during contact with the schools. Although the instrument is a self-reported questionnaire, students with LD could not complete it by themselves due to their reading deficiencies. To ensure their understanding, each questionnaire item was read aloud to the students with LD before they completed the questionnaires by themselves. Student without LD were invited to fill in the questionnaire during computer class or in the morning before classes began. They were allowed to ask questions as they filled in the questionnaire, including the meaning of items.

**Data analysis**

The initial ICT competency measurement model, in which all the coefficients among factors were open for estimation, was tested and analyzed by structural equation modeling (SEM) using the software AMOS (version 22). The commonly recommended fit indices when reporting SEM analyses were conducted, including preliminary fit criteria, overall model fit, and the fit of the internal structure of the model. The preliminary fit criteria include: positive error variances, significant error variances, correlations not exceeding 1, and factor loadings ranging from .05 to .95 (Bagozzi & Yi, 1988).

For this study, several indices were used to test the overall model’s fit, including the goodness-of-fit index (GFI), adjusted goodness-of-fit index (AGFI), comparative fit index (CFI), root mean square error of approximation (RMSEA), standardized root mean residuals (SRMR), normed-fit index (NFI), and parsimonious normed-fit index (PNFI). For goodness of fit to be demonstrated, the recommended values for these indices should be as follows: GFI > .90, AGFI > .90, RMSEA < .08, SRMR < .05, NFI > .90, and PNFI > .50.

In addition to the indices regarding the overall model’s fit, reliability and convergent validity of the factors within the proposed model were estimated, using composite reliability (CR) and average variance extracted (AVE) were used. The criteria for CR is > .60 and AVE is > .5.
First, the authors tested the fitness of the initial model using the above indices. Meanwhile, the PNFI was used to determine if the initial model was parsimonious. Factors with non-significant direct effects on predicted variables were excluded if they negatively impacted the fitness of the initial model. The revised model that resulted from the initial model was used to answer the abovementioned hypotheses.

Meanwhile, multiple-group analysis was used to test the measurement invariance between the elementary school junior high school models to explore cross-validation. $\Delta \chi^2$, $\Delta$NFI, $\Delta$IFI, $\Delta$RFI and $\Delta$TLI between the two models were included as indicators to represent invariance of measurement weights, structural weights, structural covariance, structural residuals, and measurement residuals. Measurement invariance would be accepted if $\Delta \chi^2$ was insignificant ($p \geq .05$) and the absolute value of $\Delta$NFI, $\Delta$IFI, $\Delta$RFI and $\Delta$TLI were lower than .05 (Little, 1997). The two models would be regarded as having no significant differences based on acceptance of the measurement invariance criteria.

**Results**

**An SEM model for ICT competencies**

Pearson product-moment correlation was adopted to test the relationship among the explored variables. As the correlation coefficients (shown on Table 2) suggest, each variable has a significant relationship with at least one of the others variables. As such, all variables were included in the full model.

### Table 2. Correlation matrix

<table>
<thead>
<tr>
<th></th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.Disability</td>
<td>0.00</td>
<td>0.17</td>
<td>-0.318**</td>
<td>-0.146</td>
<td>-0.225**</td>
<td>-0.061</td>
<td>-0.236**</td>
<td>-0.166*</td>
<td>-0.397**</td>
</tr>
<tr>
<td>2.Grade</td>
<td>-</td>
<td>0.01</td>
<td>0.171**</td>
<td>-0.003</td>
<td>0.080</td>
<td>-0.026</td>
<td>-0.113</td>
<td>0.039</td>
<td>0.193**</td>
</tr>
<tr>
<td>3.Family ownership of a computer</td>
<td>-</td>
<td>-</td>
<td>0.090</td>
<td>0.143*</td>
<td>0.260**</td>
<td>0.195**</td>
<td>0.194**</td>
<td>0.147*</td>
<td>0.162*</td>
</tr>
<tr>
<td>4.Office software use</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.363***</td>
<td>0.467**</td>
<td>0.284**</td>
<td>0.410**</td>
<td>0.282**</td>
<td>0.530**</td>
</tr>
<tr>
<td>5.ICT for learning</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.238**</td>
<td>0.160*</td>
<td>0.359**</td>
<td>0.152*</td>
<td>0.201**</td>
</tr>
<tr>
<td>6.ICT for social</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.520**</td>
<td>0.462**</td>
<td>0.313**</td>
<td>0.572**</td>
</tr>
<tr>
<td>7.ICT for leisure</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.453**</td>
<td>0.274**</td>
<td>0.329**</td>
</tr>
<tr>
<td>8.ICT for life needs</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.210**</td>
<td>0.370**</td>
</tr>
<tr>
<td>9.ICT attitude</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10.ICT competency</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*Note: *$p < .05$; **$p < .01$.**

### Table 3. Standardized path coefficients of the initial and revised ICT competence models

<table>
<thead>
<tr>
<th>Model</th>
<th>Independent variable</th>
<th>Outcome variable</th>
<th>Direct effect</th>
<th>$p$</th>
<th>% of variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>Family ownership of a computer</td>
<td>ICT Attitude</td>
<td>0.11</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Office software use</td>
<td>ICT Attitude</td>
<td>0.19*</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Grade</td>
<td>ICT Attitude</td>
<td>-0.03</td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Disability</td>
<td>ICT Attitude</td>
<td>-0.07</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ICT for social</td>
<td>ICT Attitude</td>
<td>0.13</td>
<td>0.17</td>
<td>23%</td>
</tr>
<tr>
<td></td>
<td>ICT for life needs</td>
<td>ICT Attitude</td>
<td>-0.09</td>
<td>0.35</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ICT for leisure</td>
<td>ICT Attitude</td>
<td>0.25**</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ICT for learning</td>
<td>ICT Attitude</td>
<td>0.06</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ICT Attitude</td>
<td>ICT Competencies</td>
<td>0.37***</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Family ownership of a computer</td>
<td>ICT Competencies</td>
<td>0.01</td>
<td>0.85</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Office software use</td>
<td>ICT Competencies</td>
<td>0.15*</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Grade</td>
<td>ICT Competencies</td>
<td>0.15**</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Disability</td>
<td>ICT Competencies</td>
<td>-0.26***</td>
<td>0.00</td>
<td>67%</td>
</tr>
<tr>
<td></td>
<td>ICT for social</td>
<td>ICT Competencies</td>
<td>0.33***</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ICT for learning</td>
<td>ICT Competencies</td>
<td>-0.04</td>
<td>0.47</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ICT for leisure</td>
<td>ICT Competencies</td>
<td>-0.03</td>
<td>0.63</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ICT for life needs</td>
<td>ICT Competencies</td>
<td>0.08</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>Revised</td>
<td>ICT for leisure</td>
<td>ICT Attitude</td>
<td>0.30***</td>
<td>0.00</td>
<td>20%</td>
</tr>
<tr>
<td></td>
<td>Office software use</td>
<td>ICT Attitude</td>
<td>0.25***</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>
ICT Attitude I ICT Competencies I .38*** I .00
ICT for social I ICT Competencies I .37*** I .00
Office software usage I ICT Competencies I .15* I .02 65%
Grade I ICT Competencies I .15** I .00
Disability I ICT Competencies I -.28*** I .00

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

SEM was adopted to test the proposed initial measurement model. The results obtained from the analysis indicated a good overall fit of the initial model ($\chi^2 = 65.951, p < .001$; GFI = .959; AGFI = .866; CFI = .960, RMSEA = .076; SRMR = .028; NFI = .936) and the internal structure of the model was also consistent (ICT attitude CR = .6874, AVE = .5284; ICT competencies CR = .8670, AVE = .6855). This theoretical model explained 67% of the total variance for ICT competency.

In this model, the purpose of using ICT for leisure activities and frequency of Office software use had positive direct effects on attitude towards ICT (.30, $p = .00$; .25, $p < .00$). These two model explained 20% of the total variance for attitude towards ICT (shown in Table 3).

The most powerful direct positive effect on ICT competency was ICT attitude (.37, $p = .00$). The results also indicate that grade, the purpose of using ICT for social interaction, and frequency of Office software use had direct positive predictive effects (.15, $p = .01$; .33, $p = .00$; .15, $p = .02$) while disability had a direct negative effect (-.26, $p = .00$) on ICT competency (shown in Table 3).

This initial model also explained 23% of the total variance for attitude towards ICT. The results indicate that frequency of Office software use and the purposes of using ICT for leisure activities were positively and significantly associated with attitude towards ICT (.19, $p = .03$; .25, $p = .01$) (shown in Table 3).

![Figure 2. The standardized path coefficients of the revised ICT competency model: Elementary students sample](image)

Although most of the indices of the proposed initial model met the goodness-of-fit criteria, the PNFI index, indicating the parsimony of the model, was poor (PNFI = .336), meaning the model should be simplified. Furthermore, the lack of significant path coefficients also illustrated the existence of many unrelated factors. Therefore, a new revised ICT competency model (Figure 2) was proposed by excluding insignificant paths.
The new revised model explained 65% of the total variance of ICT competency with better goodness of fit for the elementary student sample ($\chi^2 = 60.621, p < .001$; GFI = .951; AGFI = .901; CFI = .958; RMSEA = .073; SRMR = .0392; NFI = .928; PNFI = .557; ICT attitude CR = .6791, AVE = .5169; ICT Competencies CR = .8581, AVE = .6691).

In this revised model, the factors which have direct and significant effects on ICT competency are disability, grade, using ICT for social interaction, frequency of Office software use, and attitude towards ICT. The attitude towards ICT latent variable also has a direct positive effect on ICT competency (.38, $p = .00$). Disability, grade, frequency of Office software use and using ICT for social purpose also have the direct predictive effects on ICT competencies (.28, $p = .00$; .15, $p = .00$; .15, $p = .02$; .37, $p = .00$) (shown in Table 3).

Overview of the hypotheses

Based on the significance of the standardized path coefficients shown in Table 2, the hypothesis testing results are summarized in Table 4. H1 (1a, 1b), H3 (3a, 3b), and H5 (5a, 5b, 5c, 5d) were proposed to test the effects of demographics, ICT access, and purposes for using ICT on attitude towards ICT. As Table 4 shows, only two hypotheses (H3b, H5c) were accepted, which suggests that the frequency of using Office software and the use of ICT for leisure activities do have positive effects on ICT attitude. Family ownership of a computer, grade, disability, and using ICT for learning, social interaction, and daily life needs had no effects on ICT attitude.

<table>
<thead>
<tr>
<th>Hypotheses</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesis 1a: Students without LD will have a better attitude towards using ICT than students with LD.</td>
<td>Rejected</td>
</tr>
<tr>
<td>Hypothesis 1b: Higher grades will be associated with more positive ICT attitudes.</td>
<td>Rejected</td>
</tr>
<tr>
<td>Hypothesis 2a: Students without LD will have higher levels of ICT competency than students with LD.</td>
<td>Accepted</td>
</tr>
<tr>
<td>Hypothesis 2b: Higher grades will be associated with higher levels of ICT competency.</td>
<td>Accepted</td>
</tr>
<tr>
<td>Hypothesis 3a: Students who have a computer at home will have more positive ICT attitudes than students without a computer at home.</td>
<td>Rejected</td>
</tr>
<tr>
<td>Hypothesis 3b: The frequency of using Office software will be associated with more positive ICT attitudes.</td>
<td>Accepted</td>
</tr>
<tr>
<td>Hypothesis 4a: Students who have a computer at home will have higher levels of ICT competency.</td>
<td>Rejected</td>
</tr>
<tr>
<td>Hypothesis 4b: The frequency of using Office software will be associated with higher levels of ICT competency.</td>
<td>Accepted</td>
</tr>
<tr>
<td>Hypothesis 5a: Higher levels of using ICT for learning activities will be associated with more positive ICT attitudes.</td>
<td>Rejected</td>
</tr>
<tr>
<td>Hypothesis 5b: Higher levels of using ICT for social interaction will be associated with more positive ICT attitudes.</td>
<td>Rejected</td>
</tr>
<tr>
<td>Hypothesis 5c: Higher levels of using ICT for leisure activities will be associated with more positive ICT attitudes.</td>
<td>Accepted</td>
</tr>
<tr>
<td>Hypothesis 5d: Higher levels of using ICT for life activities will be associated with more positive ICT attitudes.</td>
<td>Rejected</td>
</tr>
<tr>
<td>Hypothesis 6a: Higher levels of using ICT for learning activities will be associated with higher ICT competency.</td>
<td>Rejected</td>
</tr>
<tr>
<td>Hypothesis 6b: Higher levels of using ICT for social interaction will be associated with higher ICT competency.</td>
<td>Accepted</td>
</tr>
<tr>
<td>Hypothesis 6c: Higher levels of using ICT for leisure activities will be associated with higher ICT competency.</td>
<td>Rejected</td>
</tr>
<tr>
<td>Hypothesis 6d: Higher levels of using ICT for daily life activities will be associated with higher ICT competency.</td>
<td>Rejected</td>
</tr>
<tr>
<td>Hypothesis 7: Students’ with more positive ICT attitudes will demonstrate higher levels of ICT competency.</td>
<td>Accepted</td>
</tr>
</tbody>
</table>

H2 (2a, 2b), H4 (4a, 4b), H6 (6a, 6b, 6c, 6d), and H7 were proposed to examine the effects of demographics, ICT access, purposes for using ICT, and attitude towards ICT on ICT competency. As the results shown in Table 4, five hypotheses were accepted (H2a, H2b, H4b, H6b, H7). Disability, grade, frequency of Office software use, using ICT for social activities, and ICT attitude had significant effects on ICT competency. However,
family ownership of a computer and using ICT for learning, leisure, and daily life needs had no effects on ICT competency.

Examining the revised model with a different sample

Data collected from junior high school-age students were used to examine the goodness of fit of the revised model generated from data from elementary school-age students. The revised model (assuming the unconstrained model is correct) had goodness of fit ($\chi^2 = 132.042$, $p < .001$; GFI = .944; AGFI = .885; CFI = .944, RMSEA = .058; SRMR = .0390; NFI = .912; PNFI = .547; attitude towards ICT: CR = .6352, AVE = .4688; ICT Competency: CR = .8246, AVE = .6119). The model explained 64% of the total variance of ICT competency. Meanwhile, it also explained 27% of the total variance of attitude towards ICT.

In this revised model for junior high school students, the attitude towards using ICT was able to predict ICT competency, with a direct and significant effect (.49, $p < .001$). The factors which have direct effects on ICT competency are disability, the purpose for social interaction, and frequency of Office software use (-.31, $p < .001$; -.27, $p < .001$; .15, $p = .04$). These results also indicate that the purposes of using a computer for leisure and frequency of Office software use have direct predictive effects on ICT attitude (.35, $p < .001$; .32, $p = .04$). Furthermore, the results of the measurement invariance test also indicate that these two ICT competence models - elementary school and junior high school - was invariant ($\Delta \chi^2 = 1.436$, $p = .488$; $\Delta$NFI = -.001; $\Delta$IFI = 0; $\Delta$RFI = -.003; $\Delta$TLI = -.003). The results also reveal that the measurement weights, structural weights, structural covariance, structural residuals, and measurement residuals between these two models were equal, which indicate that no significant difference between the models, thereby confirming the cross-validation of the revised model.

Discussion

The present study adopts SEM to explore the roles of proposed variables in explaining students’ ICT competency. The revised measurement models shown in Figure 2 and 3 are able to address our research questions. The results of SEM show that the initial ICT competency model had goodness of fit but poor parsimony. Therefore, a revised ICT competency model with parsimony and goodness of fit was generated from

---

**Figure 3.** The standardized path coefficients of the revised ICT competency model: junior high school students

Chi-square = 132.042 (df=54) $p<.001$
GFI= .944 AGFI= .885 CFI= .944
SRMR= .0390 RMSEA= .058
NFI= .912 PNFI= .547
ICT Attitude CR= .6352 AVE= .4688
ICT Competencies CR= .8246 AVE= .6119
data from elementary school age students. Furthermore, this new model was also able to help explain the complex phenomena of ICT competency for junior high school age students, as there was no difference between two models in terms of elementary school and junior high school students.

The explanatory variables, ICT attitude, disability, grade, frequency of Office software use, and using ICT for social and leisure purposes, could explain the 65% and 64% of total variance of ICT skills for elementary and junior high school age students. The percentage of variance explained was as high as the initial ICT competency model (67%). Meanwhile, using ICT for purposes of leisure and frequency of Office software use have direct effects on attitude towards ICT. The other three proposed variables, (1) family ownership of a computer, (2) using ICT for purpose of learning activities and (3) daily life activities, were excluded from the revised model because of their insignificant effect on ICT competence or attitude towards ICT.

One of the most important variables was that of attitude towards using ICT which demonstrated a positive and significant path coefficient with ICT competency, meaning that higher motivation and confidence towards using ICT may influence ICT competency. This study also supports Martin’s (2003) perspective, that motivation should be regarded as an important factor for using ICT. It provides a new perspective on understanding the digital divide. Instead of looking at demographic factors, finding strategies for encouraging students to use ICT in their daily life and for learning activities could be a key to enhancing their ICT skills and competencies.

It is noteworthy that frequency of Office software use was the only a factor that influenced both ICT competency and attitude towards ICT. Students with greater frequency of Office software use (such as Word and Power Point) demonstrated better ICT competency and had a more positive attitude towards using ICT. One reason could be that the more one uses ICT, the better competency one develops. This relationship could also be explained by the direct positive effect of using ICT for purpose of social interaction on ICT competency. In addition, the level of using ICT for social interaction reflects the frequency of using social media. Meanwhile, the competency of using social media might be related to ICT competency. For example, text entry is essential for using social media; uploading pictures or videos is also popular when using social media.

For both elementary to junior high school students, using ICT for learning is an important issue. However, the abovementioned revised model does not include the factor of using ICT for learning. One of the possible reasons might be the low reported use of ICT for leaning activities, with an average of 1.80 for elementary students and 1.75 for junior high students on a 4-point rating scale. Because of low participation, the importance of using ICT for the purpose of learning is not fully appreciated by students. The authors believe that using ICT for learning could provide opportunities to practice their ICT competency which they also learn from ICT courses in school. This perspective might also be supported by the positive effect of using ICT for social interaction on ICT competency.

From the results of this study, it appears that providing formal ICT courses in school is insufficient. Alongside ICT courses, school should consider creating learning activities which require using ICT, for example for reports, presentations, and online discussions. Additionally, a variety of studies has demonstrated the effects of using ICT for learning, especially for students with disabilities (Starcic & Bagon, 2014). By integrating ICT with learning activities, students can practice and apply the competencies they have learned in the ICT courses, thereby benefitting their ICT competency. This perspective could also be supported by the significant positive effect of the frequency of using Office on ICT competency.

Disability is the single most important issue for us to explore. As the results indicate, disability has a negative effect on ICT competency. Students with LD reported lower ICT competency than their non-LD peers for both groups. The results confirm those of previous studies (Vicente & López, 2010; Wu et al., 2014), which regard disability as an important factor in the persistence of a digital divide. The results of this investigation also provide us an opportunity to reflect on the fact that students with LD perform poorer on ICT competency, even though they take ICT courses with their non-LD peers at school. It is possible that learning ICT skills using the same materials and methods may be inappropriate for students with LD (Wu et al., 2014).

However, disability did not have a significant direct effect on attitude towards using ICT. In other words, the reported attitudes towards using ICT were not significantly different for students with and without LD for both groups. Meanwhile students with a more positive attitudes perform better in terms of ICT competency. In the future, in order to promote attitudes towards ICT, we might offer students with LD specially designed ICT courses which can meet their individual learning characteristics, e.g., difficulties with reading, text entry, and working memory (Hallahan et al., 2004). For example, by providing more steps for the operating process based on the results of task analysis or offering alternative text entry methods, such as voice recognition. We can
examine whether or not the influence of disability on ICT competency could be eliminated through offering appropriate ICT courses.

Age was regarded as one of the important factors concerning ICT competency and the digital divide in the previous studies (van Deursen et al., 2011; Enoch & Soker 2006). Grade might represent the effect of age in this study. Grade has a significant positive effect on ICT competency for the elementary school group but has no significant impact for the junior high school group. A ceiling effect might exist, since these ICT competencies which were investigated in this study included basic computer operations, Web use, and Office software use, all of which are essential but basic skills. Students have already started to learn these ICT skills from third grade in elementary school. The previous studies recruited participants ranging from teenagers to elders, but the participants of the current study were all digital natives (e.g., van Deursen et al., 2011). Therefore, the effect of grade on ICT competency was weak for the elementary school group and had no effect for the junior high school group. Meanwhile, we are unsure whether these differences will endure into the future.

The revised model also supports the notion that the digital divide has shifted from physical access to competency and usage (van Dijk, 2006). The “haves and have-nots” (Warschauer, 2003) is not a factor in the model, since more than 90% of students reported having one or more computers at home. Furthermore, the frequency of software usage, and attitude towards using ICT should be emphasized.

The present study illustrates that a revised model, derived through SEM, can illustrate structural relations among the proposed factors and ICT competency in a holistic and comprehensive framework. The model reflects the current situation only. As mentioned above, specially designed ICT courses for students with LD should be provided. And the effect of a new factor, ICT training programs, on ICT competency and attitude towards using ICT should be explored in the future.

In addition, the current study focused on recruiting students with learning disabilities. Although learning disabilities account for the highest percentage of school-age students with disabilities (Hallahan et al., 2004), different types of disabilities might have demonstrated different relationships among the same variables. Therefore, future studies could recruit students with other types of disabilities to examine whether or not the model still fits.

The ICT competency defined in this study focused on traditional laptops and desktops only. But the diffusion of handheld devices, e.g., smartphones and tablets, has reached high levels in developed and developing countries. Meanwhile, mobile learning or ubiquitous learning is also regarded as an essential trend for teaching and learning. Future studies could explore competencies relevant to the operation of handheld devices.

References


Mediating Effects of Individuals’ Ability Levels on the Relationship of Reflective-Impulsive Cognitive Style and Item Response Time in CAT

Chao Wang and Hong Lu*

Department of Educational Technology, Shandong Normal University, China // wxcstudent@126.com // sdnulbh@163.com

*Corresponding author

(Submitted December 9, 2016; Revised April 13, 2017; Accepted April 17, 2017)

ABSTRACT

This study focused on the effect of examinees’ ability levels on the relationship between Reflective-Impulsive (RI) cognitive style and item response time in computerized adaptive testing (CAT). The total of 56 students majoring in Educational Technology from Shandong Normal University participated in this study, and their RI cognitive styles were diagnosed using the Matching Familiar Figures Test-20 (MFFT-20). Examinees’ ability values and average item response time were recorded by the computerized adaptive testing system. Then mediation analysis was implemented and the findings revealed that there was direct and indirect effect between RI cognitive style and item response time in CAT. What’s more, RI cognitive style also directly affected the ability levels, and then the ability levels impacted on item response time. So, examinee’s ability level was a partly mediator between RI cognitive style and item response time. Furthermore, RI cognitive style of the examinees might also be diagnosed according to ability values and average item response time. The relevant research and implications were further discussed.

Keywords
Computerized adaptive testing, Reflective-Impulsive cognitive style, Ability level, Item response time, Mediation

Introduction

Compared with the traditional linear tests, each examinee in computerized adaptive testing (CAT) takes a unique test that is dynamically adjusted to his/her performance level. After each response in CAT, the ability estimate is updated and the next item is selected such that it has optimal properties according to the new estimate (van der Linden & Glas, 2003). So the item administration would be neither too hard nor too easy for each examinee to keep them appropriately challenged and more likely to stay engaged (Green, 1983; Wainer, 2000). Additionally, by selecting items of particular relevance to an individual response, CAT applications can concurrently reduce the number of questions, increase measurement precision, decrease response burden, and cover a wide measurement range (Devine et al., 2016). It also takes less time to complete the gain of the accurate results by using more efficient and precise assessments. Because of the above mentioned advantages, CAT is becoming more common in high-stake assessment. For example, in the United States, several large tests have an operational CAT version, e.g., the Graduate Record Examination (GRE) and the Computerized Placement test. Several licensure boards have also implemented CAT versions of their tests, including the National Council of State Boards of Nursing and National Board of Medical Examiners. In addition, the US Department of Defense has implemented a CAT version of the Armed Services Vocational Aptitude Battery. CAT is also becoming increasingly popular outside the US. For example, in the Netherlands, the National Institute for Educational Measurement has recently released two CATs, one for assigning examinees to different levels of a mathematics course and the other for assessing achievement in a specific mathematics course (Meijer & Nering, 1999; Verschoor & Straetmans, 2000).

Review of literature

CAT has multiple theoretical advantages over standard static assessments and is becoming an evolutionary step to future testing methodologies, therefore, there are an increasing number of researchers dedicating to it. Most of the researches on CAT were focused on the technical aspects, such as the construction of item pool (Ariel, Veldkamp, & van der Linden, 2004; Barla et al., 2010; Reckase, 2010), the judgment of initial conditions (Kozierkiewicz-Hetmańska & Nguyen, 2010; Mansoor, 2007; Wauters, Desmet, & Van den Noortgate, 2010), the strategies of item selection which included content balancing (Chen & Ankenmann, 2004; Cheng, Patton, & Shao, 2014; Su, 2016; Yi & Chang, 2003) and item exposure control (Leroux, Lopez, Hembry, & Dodd, 2013; Leung, Chang, & Hau, 2005), the estimate of ability (Balasim, 2009; Huang, Lin, & Cheng, 2009) and the terminal condition of the test (Kozierkiewicz-Hetmańska & Nguyen, 2010; Mansoor, 2007; Triantafillou, Georgiadou, & Economides, 2008). Besides the technical aspects, the psychological reactions of examinees to
CAT also had been investigated and described in detail. For example, Kim and McLean (1994) declared that test anxiety was generally found to be negatively related to test performance on the CAT. And the following researches paid attention to the question whether or not CAT would be more anxiety-producing for examinees than conventional testing and whether or not CAT could produce a disadvantage for examinees with higher anxiety. Fritts and Marszalek (2010) noted that traditional paper-and-pencil tests (P&P) examinees had higher anxiety than CAT examinees when controlling for trait test anxiety and computer anxiety. Some researchers also paid attention to the potential of CAT to activate examinees’ motivations (Lu, Hu, Gao, & Kinshuk, 2016; Ortner, Weißkopf, & Koch, 2014; Rheinberg, Vollmeyer, & Burns, 2001; Tonidandel, Quiñones, & Adams, 2002). Compared with large number of researches which focused mainly on the technical aspect and psychological effects of CAT, a little attention was paid to item response time of different examinees in CAT.

CAT provides not only examinees’ ability values but also item response time. Typically, in the test, different examinees have different item response time in different items. It is therefore particularly interesting to investigate if there are differences of item response time between different conditions and what additional meaning can be gained from item response time. The evidences that incorrect answers took more time than correct were often reported. For example, Hornke (2000) tested 5912 young men with a computerized adaptive test and showed that item response time for wrong solutions was noticeably longer than for correct solutions. The reason of this difference was also be expounded by Hornke (2005). He indicated that the correct answer seemed to “catch the eye” of the examinee, while wrong answers were rather the result of a longer process. Perhaps item details were repeatedly considered, then discarded, and finally forgotten. The effort to find the answer dragged on, and in the end it might be terminated by a random guess. Chang, Plake and Ferdous (2005) investigated time demands on a time restricted, fixed-length CAT and found that across item blocks, high ability examinees spent more time on test questions than did low ability examinees. Beside this difference between higher and low ability examinees, another difference was that high ability examinees spent more time averaged across items they answered incorrectly than ones they answered correctly, but low ability examinees systematically spent nearly equivalent time on items that they answered correctly and incorrectly. Some studies demonstrated that item response time, on average, increased with the item difficulty (Bergstrom, Gershon, & Lunz, 1994; Bridgeman & Cline, 2000). The adaptive algorithm often provided more difficult items for examinee who had high level of competency. So, high ability examinees received more time-consuming items. However, Chang, Plake and Ferdous (2005) employed pretest items which were not tailored to the examinees’ ability level to verify that viewpoint and found that regardless of whether the examinees answered the items correctly or incorrectly, the fact that high ability people spent more time on all items was not necessarily related to the difficulty of the item. And they suggested that high ability examinees might have a higher persistence with test questions.

Item response time is considered the most important measure used to investigate the hypotheses about mental processing in cognitive psychology (Eysenck & Keane, 2010). In CAT, examinees are administered test questions that are matched to their ability levels, so this item selection algorithm would create a challenging and optimally motivating assessment situation in which examinees feel neither over- nor under-challenged. Therefore, it’s reasonable to assume that in CAT, the difference of item response time among the examinees should also be reflected by some elements of cognitive psychology, such as the Reflective-Impulsive (RI) cognitive style, which is one of the most frequently studied cognitive styles and is related with the speed of solving problems. Hence, the present study attempts to verify the direct and indirect effect of RI cognitive style on item response time.

Theoretical foundations

Everyone has different approaches and propensities of how to learn new things. Some people prefer to learn from theoretical information while others rather tend to learn from practical experience. The above-mentioned difference is embodied in the implication of cognitive style in the field of psychology. Cognitive style refers to individual differences when perceiving, attending, remembering, deciding, and solving problems (Quiroga, Martínez-Molina, Lozano, & Santacreu, 2011). RI cognitive style, first proposed by Kagan, Rosman, Day, Albert and Phillips (1964) using the Matching Familiar Figures Test (MFFT), is one of the most constantly researched cognitive styles. When solving problems, there are many situations where students have to make decisions under great uncertainty. RI cognitive style is defined as a property of the cognitive system that refers to individual differences of response speed and accuracy in the situations of individuals’ information-processing, hypothesis-constructing, decision-making as well as problem-solving with response uncertainty (Kagan & Kogan, 1970).
Reflective individuals employ an analytic processing mode, while impulsive people apply a holistic processing mode (Ancillotti, 1984, 1985; Quiroga, Martínez-Molina, Lozano, & Santacreu, 2011). Reflection means a delay in response latency until being sure about the correct choice. Accordingly, reflective individuals show higher accuracy but higher response latencies. On the contrary, impulsivity implies a quicker choice of a response alternative. Impulsive individuals tend to accept the first hypothesis that emerges in their minds without checking its accuracy, therefore, they show lower response latencies but greater inaccuracy (Quiroga, Martínez-Molina, Lozano, & Santacreu, 2011).

The present study

A short summary that impulsive individuals tended to response much faster than reflective ones but make more mistakes in their decisions could be got from theoretical foundations section. Accordingly, Hypothesis 1 and Hypothesis 2 were as following.

**H1.** The difference of item response time between reflective and impulsive individuals was significant in CAT.

**H2.** The difference of ability levels between reflective and impulsive individuals was significant in CAT.

From the section of research review, a conclusion that high ability examinees spent more time on the items than did low ability examinees in CAT could be drawn. In other words, the examinees’ abilities would influence their item response time in CAT. Therefore, this study constructed Hypothesis 3.

**H3.** The difference of item response time between different ability individuals was significant in CAT.

According to the above three assumptions, the fourth hypothesis was presented as following.

**H4.** The ability of the examinee had a mediating effect between RI cognitive style and item response time in CAT.

The model of the mediating effect in this present study was shown in Figure 1.

![Figure 1. Ability level as mediator of RI cognitive style and item response time](image)

a: The regression coefficient was obtained by regressing the ability values on RI cognitive style. 
b: The regression coefficient between item response time and ability levels was obtained by regressing the average item response time on RI cognitive style and the ability values. 
c': The regression coefficient between item response time and RI cognitive style was obtained by regressing the average item response time on RI cognitive style and the ability values.

**Methodology**

**Participants and procedure**

The procedure of the present study was divided into three phases. The first phase was to develop a computerized adaptive testing system which could record the ability values as well as the average item response time of each examinee. The ability value was a good indicator of examinee’s ability level and the average item response time was used as the metric of item response time. In order to create the item bank of the adaptive testing, five professors developed 180 multiple choice items in respect of the probability theory and mathematical statistics. The 180 items were divided into four parallel tests. Each test was composed of 40 independent items and 20 anchor items. During the course of confirming the item parameters based on item response theory (IRT), 3524 students from four Chinese universities in Shandong province were recruited to take the pretest. Of course, the students had taken the probability theory and mathematical statistics as compulsory course. In each university, the participants were randomly separated into four groups and each group was assigned one test. On the basis of the students’ responses, the parameters of each item were calculated by employing the Bayesian Expected A
Posterior method with BILOGMG 3.0. The consequences turned out that 22 items failed to fit the two parameter logistic model, which was one of the most widely used IRT models, and were deleted in terms of the value of chi-square value and the degree of freedom. The parameters of the remaining 158 items were connected on the same scale by the mean and sigma method. Eventually, 158 items that varied according to their difficulty and discrimination constituted the item bank of the adaptive testing system.

The second phase was measuring RI cognitive style of 73 sophomore students who majored in Educational Technology in Shandong Normal University, and had taken the course of probability theory and mathematical statistics. The average age of the students was 19.09, and SD was 0.90. RI cognitive style was measured in the students using the Matching Familiar Figures Test-20 (MFFT-20; Cairns & Cammock, 1978) presented on the computer screen in this study. Before the test, a notebook computer installed with MFFT-20 was got ready. During the process of the test, each student participated in the test separately under the supervision of the experimenter. Specifically, after one examinee finished the test and went out, next one could enter the laboratory and start the test. The reason why the examinees took the test separately was to ensure the testing process of each examinee without other students’ influence and lead to more accurate data. The purposes of this phase was to choose the reflective and impulsive individuals among the 73 students, and the result was that there were 29 reflective students, 27 impulsive students, 8 fast-accurate students and 9 slow-inaccurate students. Ultimately, the results of this phase showed that the examinees who would participate in the third phase were 29 reflective and 27 impulsive students.

The third phase was the implementation of CAT. 56 students (19 male, 37 female) whose cognitive style was either reflection or impulsivity took part in the adaptive testing. CAT was conducted on a notebook computer in a quiet laboratory at Shandong Normal University. The students were tested separately just as the test process of the second phase for more accurate data. Experimental instruction was presented via the screen of the notebook computer. And the instruction explained to the examinees that they were about to take a fixed-length adaptive testing which designed to evaluate their knowledge of probability theory and mathematical statistics, and the test comprised 25 multiple choice items. After the instruction was presented, the examinees had 75 minutes to take the adaptive test, which was sufficient for them to finish the testing. At the end of the testing, the examinees would receive their ability values regarding their actual level of performance on the adaptive testing. In addition, the average item response time would also be recorded by the computerized adaptive testing system and it could only be viewed by the experimenters.

Measurement development

Reflective-Impulsive cognitive style

The examinees were asked to complete the MFFT-20 developed by Cairns and Cammock (1978), and in the present study the test was in the form of stand-alone version accomplished by the researchers. The feasibility of computer controlled MFFT-20 administration was demonstrated by van Merriënboer and Jelsma (1988). Twenty stimulus items were comprised in the MFFT-20, which was suitable for use with children and youths in the age range 7–21 yrs. Each stimulus item consisted of a standard figure and eight alternatives, and it required a visual match of the alternatives to the standard. Only one of the alternatives was identical to the standard. That was, the examinees were required to choose the alternative that was identical to the standard figure. If they chose wrongly, the borders of computer screen would flash red three times and the message would appear on the information window to require the examinee to try again until the right one was clicked. The total number of errors on the test would be collected and the average time to first response would be calculated. A double median-split method was employed to classify individuals according to the average time to first response and the total number of errors (Kagan, 1966a; Kagan, 1966b; Kagan, Pearson, & Welch, 1966). Reflective individuals had an average time to first response score above the median of the sample and a total errors score below the median of sample. In contrast, impulsive individuals had an average time to first response score below the median of the sample and a total errors score above the median of the sample. There were still another two types of individuals, and they were fast-accurate and slow-inaccurate. The former had an average time to first response score below the median of sample and a total errors score below the median of the sample. The latter one had an average time to first response score above the median of the sample and a total errors score above the median of the sample. The average time to first response as well as the total number of errors of each examinee was recorded in the data base. After the whole test was completed, the researchers figured out the median of the average time to first response score and the median of the total errors score of the sample. Then, RI cognitive style was distinguished with the above mentioned double median-split method.
The average item response time in CAT

The average item response time used as the metric of item response time meant the average time from stimulus onset to answer execution by an examinee completing a test item. During the process of adaptive testing, each examinee’s response time on each item was recorded in the database table. The system was designed to calculate the average item response time of each examinee and record it in the database as long as the examinee finished the adaptive testing.

The ability values obtained in CAT

The ability value recorded by the system was a good indicator of examinee’s ability level. In CAT, on the basis of IRT, the ability values depended on which items were responded correctly rather than on the number of items responded correctly. That was to say, the ability values of the examinees would be greater if they answered more difficult questions correctly compared with easier questions. In this study, the ability value was finally presented on the range of -3 to 3.

Statistical analysis

SPSS version 20.0 (Chicago, IL, USA) was used for analyzing the data. Significance test was done using independent-sample T test to verify the first three hypotheses. The statistical significance was set at $p < .05$. Then, linear regression analysis was performed to test the mediating effect.

Results

The difference of the average item response time between reflective and impulsive examinees

The mean of the average item response time of the examinees was 44.09 seconds, and the SD was 15.77. According to the theoretical foundations section, it was known that the reflective and impulsive individuals needed different time to answer the questions. In this study, the mean of the average item response time of the reflective examinees was 50.39 seconds, and the SD was 15.04. The mean of the average item response time of the impulsive examinees was 37.32, and the SD was 13.81. There was statistically significant difference in the average item response time between reflective and impulsive examinees ($t = 3.38$, $df = 54$, $p = .010$). That was, the average item response time of the reflective examinees were significantly longer than the impulsive examinees’.

The difference of the ability values between reflective and impulsive examinees

The ability values received by the examinees at the end of the adaptive testing represented their ability levels in the probability theory and mathematical statistics. The mean ability value of all examinees was 0.61, and the standard deviation was 0.44. Furthermore, the mean ability value of the sample of the reflection ($M = .73$, $SD = .39$) was found to be higher than the mean ability value of the sample of the impulsivity ($M = .47$, $SD = .46$). Statistically significant difference in the ability values between reflective and impulsive examinees was detected at this independent sample t test ($t = 2.34$, $df = 54$, $p = .023$). So, the ability values of the reflective examinees were significantly higher than the impulsive examinees’.

The difference of the average item response time between examinees with different ability levels

The third independent sample t test was performed to test the difference of the average item response time between examinees with different ability levels. At the end of the experiment, the examinees were divided into two groups according to their ability values. The examinee whose ability value was higher than the mean ability value of the sample was classified as high ability examinee, otherwise he or she would be classified as low ability examinee. In this study, the average time spent by each high ability examinee on each item was 51.84 seconds, and the standard deviation was 14.79. The average time spent by each low ability examinee on each item was 36.34 seconds, and the standard deviation was 12.79. There was statistically significant difference in the average item response time between high ability and low ability examinees ($t = 4.19$, $df = 54$, $p = .000$). In
other words, the average item response time of the high ability examinees was significantly longer than the low ability examinees’.

The procedure and result of mediation analysis

According to the above three results that proved H1, H2 and H3 to be true, it was meaningful to test the H4. So, whether ability level of examinee acted as a mediating variable in the relation of RI cognitive style and item response time was tested. Following the approach suggested by Wen, Chang, Hau and Liu (2004), after centering of the massive measured data, three regression analyses were conducted. The findings from the three regression analyses were summarized in Table 1, Table 2 and Table 3. From Table 1, it would be found that the first linear regression model about the average item response time on RI cognitive style (Reflection = 1, Impulsivity = 0) accounted for 17.5% ($R^2 = .175$) of the variance. And the second linear regression model about the ability values on RI cognitive style account for 9.2% of the variance. In the third multiple linear regression model, RI cognitive style and the ability values explained 39.8% ($R^2 = .398$) of the average item response time. In Table 2, the value of $F$ was the mean square regression divided by the mean square residual. The probability of the $F$-values in the three regression models showed that the probability of the given correlation occurring by chance was less than 0.05. It meant that all the three linear regression equations were significant. In Table 3, the values of $B$ were the constant and the coefficients of the linear regression equation. Beta was the $B$-value for standardized scores of the independent variables. From Table 3, it would be found that all the regression coefficients were statistically significant.

<table>
<thead>
<tr>
<th>Table 1. Results of regression analysis (Model summary)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>1$^a$</td>
</tr>
<tr>
<td>2$^b$</td>
</tr>
<tr>
<td>3$^c$</td>
</tr>
</tbody>
</table>

$^a$ = Predictors: (Constant), RI cognitive style. Dependent Variable: the average item response time.  
$^b$ = Predictors: (Constant), RI cognitive style. Dependent Variable: ability values.  
$^c$ = Predictors: (Constant), RI cognitive style, ability values.  

<table>
<thead>
<tr>
<th>Table 2. Results of regression analysis (ANOVA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>1$^a$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>2$^b$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>3$^c$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

$^a$ = Predictors: (Constant), RI cognitive style. Dependent Variable: the average item response time.  
$^b$ = Predictors: (Constant), RI cognitive style. Dependent Variable: ability values.  
$^c$ = Predictors: (Constant), RI cognitive style, ability values.  

<table>
<thead>
<tr>
<th>Table 3. Results of regression analysis (Coefficients)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>1$^a$ (Constant)</td>
</tr>
<tr>
<td>RI cognitive style</td>
</tr>
<tr>
<td>2$^b$ (Constant)</td>
</tr>
<tr>
<td>RI cognitive style</td>
</tr>
<tr>
<td>3$^c$ (Constant)</td>
</tr>
<tr>
<td>RI cognitive style</td>
</tr>
<tr>
<td>Ability values</td>
</tr>
</tbody>
</table>

$^a$ = Predictors: (Constant), RI cognitive style. Dependent Variable: the average item response time.  
$^b$ = Predictors: (Constant), RI cognitive style. Dependent Variable: ability values.  
$^c$ = Predictors: (Constant), RI cognitive style, ability values. Dependent Variable: the average item response time.
To be specific, when the average item response time was regressed on RI cognitive style without any mediator (path c in Figure 2), the coefficient was found to be significant ($c = .418, p = .001$). This result confirmed that there was a significant total effect of RI cognitive style on item response time. In the second regression, the ability values were regressed on RI cognitive style (path a in Figure 3), and the regression coefficient also arrived significance level ($a = .303, p = .023$), which revealed that RI cognitive style was significantly positively related to examinee’s ability level. Finally, multiple regression analysis about the average item response time on RI cognitive style and the ability values (path b and c’ in Figure 3) was done and both path coefficients were found to be significant ($b = .495, p = .000; c’ = .268, p = .020$). From the third regression, it could be found that the relationship of RI cognitive style and item response time was still significant when the effect of ability level was also taken into account, and it could also be found that ability level was significantly positively related to item response time when controlling for RI cognitive style. According to Wen, Chang, Hau and Liu (2004), the results of this mediation analysis implied that ability level of the examinee was a partly mediator in the relationship between RI cognitive style and item response time. The resulting model was shown in Figure 2 and Figure 3.

**Figure 2.** The relation of RI cognitive style and item response time without mediator

\[ c = 0.418^{**} \]

\textit{Note.} **$p < .01.$

$a = 0.303^{*}$

\textit{Note.} *$p < .05;$ **$p < .001.$

**Figure 3.** Mediation model showing the effect of RI cognitive style on item response time through ability level

\[ a = 0.303^{*} \]

\[ b = 0.495^{***} \]

\[ c = 0.418^{**} \]

\[ c' = 0.268^{*} \]

\textit{Note.} *$p < .05;$ **$p < .001.$

Discussion

From the results, it would be found that the RI cognitive style of examinees affected their item response time and ability levels in CAT. Reflective examinees spent more time averagely on each item but showed higher accuracy than impulsive examinees. That was, the first two hypotheses, which stated that the difference of item response time and the difference of ability levels between reflective and impulsive individuals would all be significant in CAT, were supported. The results could be explained by the implication of the RI cognitive style. What needed further explanation was that the experiment in this study was carried out with computerized adaptive testing mode. That was, the present study provided examinees with the test environment where the item administration was dynamically adjusted to the estimate of the examinee’s ability level. However, the previous studies (Haghighi, Ghanavati, & Rahimi, 2015; Spinella & Miley, 2003) were used to apply conventional linear tests with fixed item sequence, in which the examinees were forced to work on items that were either too hard or too easy for them, to detect the differences of mental characteristics between reflective and impulsive individuals. The accuracy of the results of testing the differences between reflective and impulsive individuals might be influenced by the fact that the difficulty of item was not adjusted to the examinee’s ability level in traditional linear test. In contrast to the studies with fixed item sequence testing mentioned above, the present study with computerized adaptive testing would be more scientific and persuasive.

The third hypothesis described the influence of examinees’ ability levels on item response time in CAT. This study did verify that the ability levels of the examinees influenced their response time. And this agreed with the previous researches which argued that the high ability individuals spent significantly longer time on solving problems than did low ability individuals (Chang, Plake, & Ferdous 2005; Hornke, 2000, 2005). In the interpretation of the results, Chang, Plake and Ferdous (2005) suggested that more able individuals were more
likely to persevere on test questions. However, the interpretations were just speculations which needed to be validated in future research.

To test the fourth hypothesis, the current study examined the mediating effects of ability levels on the relationship between RI cognitive style and item response time in CAT. The result of the mediation analysis indicated that RI cognitive style directly affected ability levels of the examinees, and that, ability levels then impacted on item response time. Therefore, ability level served as a mediator in the relationship between RI cognitive style and item response time in CAT, which was supportive for the fourth hypothesis. The result could be interpreted by the theoretical foundations part of this paper. It was clear that the reflective individuals used an analytic processing mode to solve problems, whereas the impulsive individuals used a holistic processing mode (Ancillotti, 1984; Ancillotti, 1985). Different methods of processing information not only made the individuals different in the response time on solving problems, but also made the individuals different in the accuracy of solving problems. Additionally, early studies had verified that individuals’ ability levels influenced the response time on solving problems (Gvozdenko & Chambers, 2006; Schnipke & Scrams, 1999). So, ability levels might be the mediator between RI cognitive style and response time when individuals solved problems. What’s more, the results also demonstrated that the effect of RI cognitive style upon item response time included both direct effect and indirect effect in CAT. And the ratio of indirect effect to direct effect was 56.0%, which could revealed that the direct effect of RI cognitive style on item response time was greater than the indirect effect.

Conclusion

This study provided novel preliminary evidence that RI cognitive style not only had direct relationship with item response time, but also affected item response time through influencing ability level of the examinee in CAT. More precisely, ability level played a partial mediation effect between RI cognitive style and item response time.

Previous studies had discussed how RI cognitive style affected response time when all the individuals faced one and the same task. However, this study investigated the effect of RI cognitive style on item response time while all the individuals resolved the test questions of which the difficulties were match to the individuals’ ability levels. Obviously, the conclusion of this study could enrich and advance the theory of RI cognitive style.

The specialists in the field of educational measurement and evaluation had been expecting the issue whether more information of the examinees could be obtained from the test results, especially their psychological traits. From the results of the present study, besides the examinees’ ability levels, the RI cognitive style of the examinees might also be diagnosed according to the ability values and average item response time recorded by the computerized adaptive testing system. On the basis of the foregoing, during the process of CAT it was reasonable to remind the examinees who belonged to impulsivity to make their decisions after careful considerations and remind the reflective examinees to pay attention to the item response time. It can not only help impulsive individuals to overcome the limitations of holistic information processing method to show their real ability levels, but also reduce the average item response time of the reflective examinees to some extent, which could improve the test economy. So, an interactive and cognitive style-friendly test environment could be established in CAT. The above discussion demonstrated that the potential of the vast information generated during a CAT could be designed to increase the precision of evaluation and to extend the utility of a test.

According to item response time of the examinees, some mathematical models or judgment criteria had been built to divided examinees' response behaviors into two sections, “rapid-guessing behavior” and “answer-giving behavior” (Chang, Plake, Kramer, & Lien, 2011; van der Linden, 2006). The results of this paper showed that the effects of RI cognitive style and ability levels on item response time were both significant. So, if the differences of RI cognitive style and ability levels could be taken into account to build different mathematical models or judgment criteria according item response time, it would be accurate and effective to distinguish “rapid-guessing behavior” from “answer-giving behavior.”

Limitation and future study

One limitation of the present study lied in sample bias. The number of participants in this study was small and their ages were mostly between 19 and 21 years. In addition, female in this sample significantly outnumbered male. These imbalances of sample might lead to universal conclusions of the study subject to certain restrictions. In order to make the conclusions of similar study be more persuasive, the size of sample should be as large as possible, and the distribution of age and gender should be more balanced in the future.
Item bank of CAT in this study was all about the probability theory and mathematical statistics. The single type of testing item wouldn’t be helpful for the universality of results. Thus in future study, another item types of CAT in experiment could be taken into consideration such as processes associated with verbal and numerical reasoning as well as perceptual discrimination tests and so on.

The success probability was set to 0.5 for the maximum-information item selection rule used in this study. However, earlier researches indicated that examinees might feel challenged by a constant success probability of $p = 0.5$ and therefore could not come to a sufficiently high answer certainly within a reasonable timeframe (Häusler, 2006; Häusler & Sommer, 2008). So, the chosen items in this study might be hard for examinees, which could influence the veracity of item response time. Häusler (2006) also found that within reasonable limits there was very little loss of information when the success probability deviated from $p = 0.5$. As a result, in the future study it was justifiable to set the success probability to 0.6 or 0.7, or use mixtures of highly informative ($p = 0.5$) and easier item ($p = 0.6$ or $p = 0.7$) in the item selection process.

Acknowledgments

This work has been funded by the Chinese National Education Science Foundation (BBA170069).

Reference


Increasing Information Reposting Behavior in Online Learning Community

Omid R. B. Speily and Ahmad A. Kardan*
Department of Computer Engineering and Information Technology, Amirkabir University of Technology, Tehran, Iran // Speily@aut.ac.ir // aakardan@aut.ac.ir
*Corresponding author

(Submitted August 31, 2016; Revised January 5, 2017; Accepted March 18, 2017)

ABSTRACT

Online learning communities (OLCs) enable their learner to access different types of information through internet based structures anywhere anytime. OLCs are among the strategies used for the production and repost of information by learners interested in a specific area to support asynchronous learning. In this respect, learners become members of a particular domain and begin posting. OLC members consist of different sites with different educational backgrounds as well as different levels of expertise. This causes the sharing of posts which may not be appropriate for different learners. It also reduces data reposting behavior and subsequently decreases participation in information sharing. Furthermore, most learners of these communities take up a lurking position toward the posts. One of the ways proposed to increase information reposts is the selection and display of effective posts for each individual. Effective posts are selected in such a way that they can be more likely to be reposted by learners based on each learner's interests, knowledge and characteristics. The present paper intends to introduce a new method for selecting k effective posts to ensure the increase of information repost and participation in OLCs. In terms of participation in OLCs, learners are divided into two groups of posters and lurkers. Some solutions are proposed to encourage lurking learners to participate in content repostings. Comprehensive evaluations indicated that the proposed method had significantly solved the presented challenges.

Keywords

Asynchronous learning, Information reposting, Lurker, Online learning community

Introduction

Conventional teaching cannot meet the increasing need of people to learn due to inadequate resources and the limitations of time and space. Therefore, individuals should find and use other methods to learn more efficiently. In addition, learning development happens in places where the learners spend most of their time (Topping, 2005). Because of their reception by internet users, online communities and social media can be used for learning (Wagner, 2011). Currently, learners not only depend on conventional learning, but also use other learning environments on internet such as online learning communities (OLCs) to increase their learning opportunities (Ke & Hoadley, 2009). OLC is a kind of highly accessible and convenient learning platform where learners can search what interests them and share knowledge beyond the restriction of time and space. OLCs are computer-supported public or private groups (social networks) on the internet that address the learning needs of their members by facilitating asynchronous learning. Learners with special expertise help other learners needing knowledge or information. Despite their diversity, all these communities follow the same process: learner posts a content and other learners repost it if they likes it. These posts are virally shared on the network. Content sharing is an essential part of OLC experience. In addition to composing posts by themselves, learners can also rebroadcast or repost other learners’ posts that they find of particular informational value. When a learner shares another learner’s post, in fact, he is participating in the development of a common knowledge in his own network range. It is not only a feature to display his favorite posts. Repost of interesting posts has an extensive effect on networks and spreads information by exposing a new audience to the content (Shi, Rui, & Whinston, 2014).

First-stage use is an important indicator of OLCs’ success, but the long-range success of OLCs lies on users’ persistent usage. Persistent-usage of OLCs can provide sufficient online learning materials and form prosperous online communication atmosphere which contribute to the long-term development of OLCs. The two fundamental interrelated challenges are access to appropriate information and shortage of participation (because of lurking behavior or low participation of some learner) in these communities. If information provided by a certain OLC can match users’ information requirement, users are satisfied easily. User satisfaction plays a significant role in his/her participation and engagement (i.e., collaboration and sharing) which is vital for the OLCs’ success. To address these challenges, this paper proposes a method to select the effective (appropriate) posts for each learner which increases the probability of repost behavior in the OLC.

Research questions addressed in this study are as follows:

ISSN 1436-4522 (online) and 1176-3647 (print). This article of the Journal of Educational Technology & Society is available under Creative Commons CC-BY-ND-NC 3.0 license (https://creativecommons.org/licenses/by-nc-nd/3.0/). For further queries, please contact Journal Editors at ets.editors@gmail.com
RQ1. How to predict effectiveness of the candidate post for each learner?
RQ2. Does the display of effective posts for various learners (lurker and poster) have an effect in increasing the reposting behavior and consequently participation of learners?

In the next section of this article, the literature on this subject are reviewed. Then, the proposed method is presented according to the nature of lurkers. The next section contains a detailed evaluation of the proposed method and the last section is devoted to summing up the study.

**Literature review**

Nowadays, online communities are extensively used and have induced fundamental changes to web-based systems. Therefore, according to the review of the performed studies, this paper is devoted to the role of online communities in learning as well as reposting behavior in online communities.

Over the last decade, with the development of web-based new application software on the basis of Web 2.0, social networks have received considerable attention. Social network services have provided this opportunity for the users to create online communities. Online communities improve interaction, information exchange, and personal experiences between users. The statistics reported in Duggan, Ellison, Lampe, Lenhart, and Madden’s (2015) study show that in recent years the use of online communities has dramatically been increased in various parts of the world. Web 2.0 environment and tools such as messaging, e-mail, forum, wiki, social networks, and web conference are used in developing OLCs (Barczyk & Duncan, 2013; Lambić, 2016). OLCs are used outside the classroom for supporting the learning process. In addition, the use of social media in learning purposes is about to expand. Social media capabilities such as the ability in making synchronous or asynchronous connections, tagging, posting, creating and organizing virtual groups, and resource management and sharing have provided the possibility of easily implementation of OLCS (Mazman & Usłuel, 2010). Learning with social media takes place with greater ease due to its reception capability and mobile nature in comparison with other platforms. Nowadays, many learners use social media for sharing information and knowledge, collaborating in conducting team projects, and discussing ideas and concepts (Dabbagh & Kitsantas, 2012). For example, it has been shown that learners used Facebook as a learning management system and it also satisfied them (Wang, Woo, Quek, Yang, & Liu, 2011) or based on (Chu & Meulemans, 2008) using social media are common between learners and 90% of information exchange between the learners happens via Facebook and MySpace.

In recent years, many issues regarding information sharing in OLCs have been proposed. According to (Deng & Tavares, 2013), information exchange has an important role in inspiring membership and activity in online communities and it also has a direct relationship with the value of an online community in the eyes of its members. In OLCs, sharing information improves the knowledge and skill of all group members. In such communities, sharing information increases the learner’s tendency to participate and engage in the learning process (Junco, 2012). Information sharing is defined as an activity through which members exchange information, experience, and skill among themselves. Online communities are suitable for supporting interaction and sharing between the learners. Many studies have measured the impact of a special social media on information sharing and reposting behavior (Kleinberg & Ligett, 2013; Osatuyi, 2013). Other studies have mostly investigated effective factors in online community information reposting. Interaction and information sharing between members is considered to be the most important activity in OLC. Therefore, studying information reposting behavior and providing a method to increase its amount are important factors in OLCs’ information sharing. Not many studies have been conducted on OLCs’ information sharing using information reposting behavior. The study on blogsphere addresses “epidemic” interests among different blogs regarding the content cited or copied from other blogs (Adar & Adamic, 2005). By studying the cases, it estimates the relationship between two similar blogs. By relationship, it means the use of another post in the form of citation or copy. Another important point addressed in the study is the influence of a blog on another blog via a post. (Leskovec, Adamic, & Huberman, 2007) studied sharing small pieces of text (for example news) used in other articles and texts. For this purpose, a method was implemented by which the source of each piece could be specified in the network. This made it possible to study the structure of sharing in the network. The study aims to find the sources from which a post or posts are influenced. Given the existing studies and the nature of OLCs, the present article tries to present a method for displaying the best online information posts needed by learners leading to their increased participation. In this method, the importance of the displayed posts is taken into account in terms of post topic, learner interest, and information level in addition to the type of learner (poster or lurker).
Selection of \( k \) effective posts

The learner’s profile is used to present a heuristic method for selecting effective posts for each learner. The proposed method needs no basic information on the learners because of easy-to-apply membership and using online communities (Pearson, Pearson, & Green, 2007). Only through the post content propagated in online communities does this method consider selecting effective posts for each learner. The learner’s profile include the learner’s connections, posts created or reposts from each learner. Due to the lack of a standard and diversity of implementation and design, in the proposed method the focus was only on the content assessment which is common in all kind of OLCs.

In order to choose the effective posts for each learner, this paper investigates three features: (1) subject and topic similarities of the selected post with the learner’s interests, (2) The level of expertise of the author of the candidate post, (3) The novelty of the information based on the learner’s information background. According to the learner’s profile, favorite topics of the learner can be identified by the posts they create. Posts that are selected in line with the learner’s interests are more suitable for them. Apart from interests, the learner’s knowledge level differs in OLCS and each learner with different skill levels attempts to create a post. The skill level of each post differs according to the author’s expertise. Although a post can be in line with the learner’s interests, it becomes useless when it is far from the learner’s level of expertise. Another factor which is taken into account in this paper is the novelty of information. Learning content, as a form of information posts, is useful for a learner when it is new to them. Moreover, the post’s novelty differs from one learner to another. In the following section, a detailed explanation of the procedures is provided in order to measure these three factors. This section explains selection method of \( k \) effective posts in OLCs and discusses its characteristics.

Problem statement

Suppose that an OLC is implemented under a social network such as twitter. These groups are typically displayed as graphs of followers-followees. Learners follow other learners considering their interests and expertise. This directed graph is defined as \( G = (V, E) \) where \( E \) represents the relationship between learners and \( V \) represents learners. Equation \( (u, v) \in E \) shows that learner \( u \) follows learner \( v \). If \( P \) represents the total of posts created in the whole OLC, then an online social event “post” occurs when learner \( u \) creates post \( p \) at \( t \in T \) time, represented as \( post(u, p, t) \). In the same manner, when learner \( v \) shares post \( p \) by learner \( u \) at time \( t' \), “reposting” occurs which can be represented as \( repost(v, u, p, t') \). According to the definitions provided, the probability of post can be defined as a function of the probability of repost \( p: P \times V \times V \times T \rightarrow [0, 1] \). In this function, \( T \) is the temporal domain. The probable repost of \( P \) posts by any learner from \( V \) within the temporal domain \( T \) includes the values between zero and one (zero for not posting and 1 for posting).

If \( \sigma \) is taken as the selection procedure of \( k \) effective posts from candidate posts \((v, t)\) for learner \( v \) at any \( t \) time, the output \( \sigma(v, t) \) is the \( k \) effective post to learner \( v \) at time \( t \). The total candidate posts for learner \( v \) are those already created or reposted by learners set \( V \) followed by the learner \( v \). Equation (1) shows the initial set of candidate posts for \( t' < t \). From this stream of posts, duplicate posts already displayed for the learner in the previous time \( t' < t \) should be removed.

\[
\text{Candid}(v, t) = \text{initial}(c, t) - \{p \in P | \text{post}(v, p, t') \cup a(v, t')\}
\]

Heuristic method for selection of \( k \) post

We have provided a heuristic method for the selection of \( k \) effective posts considering the computational capabilities and simplicity. This procedure is designed so as to be operational in online learning environments. In this section, the proposed method is introduced.

Similarity between posts (Post-Post)

To measure the similarities between posts, the method of text similarity is applied. For text similarity, the posts of a learner are considered as a set of word collection. According to the definition, this candidate post \( (cp) \) is similar to the posts of learner \( v \) if it is related to his interests. \( P = \{p_0, p_1, ..., p_n\} \) is the collection of \( n \) posts created by learner \( v \). To determine the relationship between a candidate post \( cp \) and the topic of interest to learner \( v \), TF/IDF method and cosine of the angle between vectors of words \( cp \) and \( P \) have been used in many references. Owing to the diversity of the employed words, this method has low accuracy. For this purpose,
different topics can be categorized in the texts using Latent Dirichlet Allocation (Bolelli, Ertekin, & Giles, 2009). Each topic includes a set of words $M_{topic} = \{w_i, w_2, ..., w_l\}$ for which the probability of the occurrence of $L$ keywords is specified in the relevant topic. Assume two vectors of candidate posts $cp_i$ and the previous posts of learner $v$ ($P_v = \{p_1, p_2, ..., p_m\}$). Topic similarity is the angle between these two vectors which is measured through cosine using formula 2.

$$\text{topsim}(cp_i, P_v) = \frac{M_{topic, cp_i} \cdot M_{topic, P_v}}{\|M_{topic, cp_i}\| \cdot \|M_{topic, P_v}\|}$$

(2)

Novelty of a post

Novelty of a post is an important characteristic influencing the post’s content value. On account of this definition, post $p_i$ is novel to learner $u_i$ if it is related to the interests of $u_i$ while it is unknown to him. Methods have been introduced in (Liu, Ma, & Yu, 2001; Padmanabhan & Tuzhilin, 1999) for the calculation of unexpected information in retrieving web pages and Gaughan and (Gaughan & Smeato, 2005) and Gabrilovich, Dumais, and Horvitz, (2004) have proposed methods for the calculation of novel news. $P_{dv} = \{p_1, p_2, ..., p_m\}$ is the set of posts written by learner $u_i$. As mentioned in the definition of novelty of a post, it is necessary to determine whether the post is related to the learner interests or not using LDA method (similar to previous section). Using LDA method, it is possible to determine the topic of each post in the form of a set of words ($M_{topic} = \{w_1', w_2', ..., w'_L\}$) with the occurrence probability of each of them in the topic. Consequently, using this method, the topic of the post is identified and the class of words related to the topic is determined. If the stop words of candidate post $cp_i$ are omitted using a standard method, and if $cp_i = \{w_1, w_2, ..., w_m\}$ is the words set used in post $cp_i$ and $M_{topic} = \{w_1', w_2', ..., w'_L\}$ is the words related to $cp_i$’s topic, then equation (3) determines the post’s novelty.

$$\text{Novelty} = \frac{\text{topsim}(cp_i, P_v)}{\text{sim}(M_{topic} - \{M_{topic} \cap cp_i\}, P_v)}$$

(3)

As seen in equation (3), as topic similarity of $cp_i$ and all $u_i$ posts (numerator of formula 3) increases, the probability of novelty of the post increases. As mentioned in the definition of post novelty, it is necessary that the topic of $cp_i$ is interesting for $u_i$. On the other hand, $M_{topic} - \{M_{topic} \cap cp_i\}$ is the set of words related to the topic of $cp_i$ which are absent in the candidate post $cp_i$. In equation (3) $\text{sim}(M_{topic} - \{M_{topic} \cap p_i\}, P_u)$ is the similarity of these words and all $u_i$ posts. When this similarity is close to 1, it indicates that $u_i$ has posted about the topic and related word of $p_i$ and is probably familiar with this post. So the probability of novelty of the post $p_i$ declines. As seen in equation (3), as the similarity of topic-related words to the learner’s topics of discussion increases, the probability of the novelty of the post declines. Post novelty is directly related to its similarity to the learner’s posts by definition.

Similarity between learners (Learner - Learner)

People with different expertise share their posts on the network. It is very beneficial to find learners with common fields. For both learners $u, v \in V$, the degree of similarity is equal to the degree of similarity between the posts already created. The essential thing about sharing information in OLCs is to find people with the same level of information in addition to similar posts. For example, a learner who has created more than 100 posts about smart phone applications is different from someone who has just had a few posts or reposts in the same field. For either learner, the action vector can be defined (Equation 4). This vector contains $n$ keywords created or reposted by the learner $v \in V$. Weighted cosine is used to determine the similarity between the vectors of learners. In this respect, the coefficients $i$ (the number of keyword repeated by learner $v \in V$) is determined for $n$ keywords till time $t$. Considering coefficients $i$, the level of learners’ knowledge on a specific area is determined according to the number of posts made by them. The two learners are examined and taken into account for determining the similarity given the repetition of the keywords in the posts.

$$\text{vector}_i = i_1 \text{keyword}_1 + ... + i_n \text{keyword}_n$$

(4)

The degree of similarity between learners $u$ and $V$ is equal to the value of cosine for the vectors of these two learners.

$$\text{sim}(u, v) = \cos(\text{vector}_u, \text{vector}_v, t) = \frac{\text{vector}_u \cdot \text{vector}_v}{\|\text{vector}_u\| \cdot \|\text{vector}_v\|}$$

(5)
In equation (5), \( \| \text{vector'} \| \) or/and \( \| \text{vector'} \| \) represent the value of action vector for learners \( v \) and \( u \), respectively. If any of these values is zero, it means that the relevant learner has had no action (neither created nor reposted). In that case, the similarity between two learners is not defined. Accordingly, in collecting data, only those learners are taken into account that have at least created 10 posts or reposted. It should be noted that the action vector of a learner changes with time. In this respect, the action vector of learners at time \( T \) is used in each determination of similarity between two learners.

### Using logistic regression to predict the probability of reposting

Logistic regression is used to estimate the probability of post effectiveness for each learner. The binary logistic model is used to estimate the probability of a binary response (reposts happened or not) based on one or more features (similarity of post-post, post novelty, and similarity between learners).

The goal of logistic regression in this problem is to find the best fitting model to describe the relationship between the effectiveness of the post (dependent variable, response or outcome variable) and a set of independent variables (3 features) using the sigmoid (logistics) function (Equation 6).

\[
y_i = h(w^T x_i) = \frac{1}{1 + \exp(-w^T x_i)}
\]

In this equation, \( y_i = 1 \) for effective posts (repost happened) and \( y_i = 0 \) for otherwise \( y_i \) is the prediction based on \( x_i \) inputs (value of each features). \( w \) is the vector of coefficients of each feature obtained from training data. In addition to classification, this method also has probable output. For example, \( h_a(x_i) = 0.8 \), means that the probability of post effectiveness is 80% for the sample \( h_a(x_i) = p(y_i=1|x_i; w^T) \).

### Probability of post effectiveness based on learner type (Lurker & Poster)

In the proposed method, in addition to the structural characteristics of learners in the network, the attention is paid to the attitude of reposting among learners. Learners are divided into two categories of posters and lurkers based on their reposting attitude (Sloep & Kester, 2009). Lurkers are those who become a member of an online educational community but do not post, are only readers, and are not active (Nonnecke & Preece, 2000; Nonnecke & Preece, 2001). Considering this type of learners, the present study intends to reduce pure lurking behavior by setting the time for online access to posts based on the importance of that post; that is, the posts which are more likely to be reposted by learners, are displayed longer for lurkers than for posters. The calculated probability of repost of post \( p \) of learner \( u \) by learner \( v \) at timestamp \( t \) (put formally \( p(\text{repost } (v, u, p, t)) \)) is shown in table (1). In algorithm (1), \( t_a \) is reposting time by learner \( u \). \( t \) is the time of running the algorithm and \( \gamma_v \) is the average time interval between the posts of learner \( v \). \( \alpha \) is the adjusted coefficient of \( \gamma_v \). At the zero line of this algorithm, by calculating the elapsed time since the repost of the posts by learner \( u \), if time is less than the average time interval between the posts of learner \( v \) \((\gamma_v)\), there is still the probability that learner \( v \) has not seen post \( p \). In this respect, the probability of reposting by learner \( v \) is equal to \( \max(p_{\gamma_v}, \alpha) \). Line 2 evaluates a condition in which the time elapsed since the repost of the posts by learner \( u \) is longer than the average time interval between the posts of learner \( v \) \((\gamma_v)\). In this case, the learner had most likely seen the post but was unwilling to repost it. Considering this fact, it is least probable that learner \( v \) reposts the post \( p \) of learner \( v \) in timestamp \( t \) \((p(\text{repost } (v, u, p, t))) \). \( \varepsilon \) is equal to minimal value at 10⁻³.

| Table 1: Algorithm: calculating the probable repost of post \( p \) of the learner \( u \) by the learner \( v \) at timestamp \( t \) |
|---|---|
| 00 | If \( t - t_a < \alpha \gamma_v \) |
| 01 | \( p(\text{repost } (v, u, p, t)) = \max(p_{u,v}, \varepsilon) \) |
| 02 | If otherwise |
| 03 | \( p(\text{repost } (v, u, p, t)) = \varepsilon \) |
| 04 | End |

The coefficient of \( \alpha \) provides more opportunity for lurking learners by adjusting the impact of \( \gamma_v \). Normally, \( \gamma_v \) of lurking learners is far longer than \( \gamma_v \) of other learners. For ease of calculation, the coefficient \( \alpha \) of \( \gamma_v \) is considered to be 1 and 2 for poster and lurker learners, respectively. In this respect, when a learner is lurking, he has more time to read and repost the posts with higher probability.
Methods

Data collection

The data presented in this study were collected based on an unrestricted self-selected web survey (Fricker, 2008). Following the methodological procedure for posting surveys on discussion boards (Ip, Barnett, Tenerowicz, & Perry, 2010), a survey invitation was published in forum with the prior approval of the online community managers. The survey was piloted between March 12 and July 12, 2015. The participants were selected from information technology and electrical engineering technophiles in popular Iranian online communities (online communities are similar to twitter environment and its members can follow other members, interested topics, etc.). Among them, only those learners are taken into accounts that had at least created 10 English posts or reposts and had more than 10 items in their list of followers and followees. In this method, qualified members were selected for a 10-day assessment period. A total of 5 to 20 posts a day were prepared for the volunteers’ assessment. Volunteers participated in the assessment independently on a daily basis without knowing other users’ opinions. These volunteers examined the posts to repost them and answered specific questions regarding each proposed feature of the posts (novelty and relevance). They ranked questions on novelty of the posts and relevance of the posts to their interests in the range of 1-5. In the 1 to 5 range of answers, 1 shows the lowest score of the feature in question, while 5 shows the highest score of that feature. Among the participants, lurkers are those who do not (re)post any posts publically in a certain period of time (Nonnecke & Preece, 2000) or those who rarely participate in the study (Sloep & Kester, 2009)’s (2009). Like the method presented in (Amichai-Hamburger et al., 2016), learners were asked about the last time they had (re)posted a post to identify lurkers. Based on their last (re)post in online communities, learners can select from more than one day, one month, one year and never. In this study, learners who selected more than one year or never were considered to be lurkers.

Sample

Non-probability sampling method was used to select more than 790 participants. In total, 608 participants participated in the study voluntarily and free of charge. Then, 583 participants were found acceptable. Therefore, 263 participants (45%) were selected as posters from 583 respondents, and 320 participants (55%) were selected as lurkers. The average age of lurkers was lower than posters; however, there was not a significant difference. According to statistics, 63% of respondents were male, and 37% were female. The majority of respondents were in 20-30 (34%) and 30-40 (41%) year old groups. The membership duration in online communities of 94% of respondents was more than 6 months; a fact which indicates the familiarity of most of them with the environment of an online community. Among them, 63% had academic educations (21% with master’s and PhD degrees) and 37% did not have academic educations. Almost 79% of them were married, and the rest were single.

Analysis

Evaluating methodology was consisted of three phases. In the first phase, for each of the participants, \( k \) posts (as the learner wished) were displayed on a daily basis. Learners decided about the post’s reposting when they observed them and they also determined when to perform the test during the day. In the second phase, according to their repost decision (previous phase), \( k \) effective and proper posts were displayed and like the first phase, learners decided about the post’s reposting. In this phase, the prediction of reposting behavior was assessed. Three well-known measures, namely precision, recall, and \( F \) measure, were used to assess the predictions (Tang, Miao, Quan, Tang, & Deng, 2015; Zhang et al., 2012). These measures are used for prediction problems of binary class (posting or not reposting). Equations 7 to 9 show these measures.

\[
\text{Precision} = \frac{|\{\text{Predicted RT}\} \cap \{\text{True RT}\}|}{|\{\text{Predicted RT}\}|} \quad (7)
\]

\[
\text{Recall} = \frac{|\{\text{Predicted RT}\} \cap \{\text{True RT}\}|}{|\{\text{True RT}\}|} \quad (8)
\]

\[
F = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (9)
\]

To evaluate the solutions for estimated relevance and novelty of a post, Kendall’s rank correlation coefficient was used to compare the measured values of both features (post novelty and relevance to the user) with the user-
expressed values (1-5 range). This method is employed to obtain the correlation between the ranks of two quantities. If \( y_i \) is the value expressed for each feature (novelty and relevance) by the user, and \( x_i \) is the proposed measured value by the proposed solution and \( (x_i, y_i) \), \( (x_j, y_j) \), ..., \( (x_n, y_n) \), then the values obtained for all of the \( n \) post are considered. Each \( (x_i, y_i) \) and \( (x_j, y_j) \) pair \( (i \neq j) \) is known as a concordant pair if \( x_i > x_j \) then \( y_i > y_j \) or if \( x_i < x_j \) then \( y_i < y_j \). The pair is discordant if \( x_i > x_j \) then \( y_i < y_j \) or if \( x_i < x_j \) then \( y_i > y_j \). If the two quantities are equal, the pair is neither concordant nor discordant. Kendall’s correlation coefficient is calculated by equation (10), and varies between 1 and -1 (-1 ≤ τ ≤ 1). If the rank of a post based on the aforementioned two features is close to its rank based on user opinions (in the questionnaire), this coefficient will be closer to 1. If the ranks differ completely, this coefficient will be closer to -1, whereas 0 shows independence of the two values.

\[
\tau = \frac{\text{number of concordant pairs} - \text{number of discordant pairs}}{\frac{1}{2}n(n-1)}
\]

(10)

The correlation of these features with repost behavior is studied in this paper. If the values of these features are significantly related to repost behavior, the relationship can be measured based on conventional learning. To this end, a method similar to Pearson’s correlation method is employed. Since each feature has continuous values and repost behavior is a binary variable (0 or 1), Pearson’s method cannot be used. Point-biserial method is utilized for such problems. If the values of each feature of each post are a continuous variable \( x \), and repost behavior for the post is a binary variable \( y \), then the point-biserial correlation coefficient is based on equation (11), where \( M_x \) shows mean feature value for posts, leading to repost behavior \( y = 1 \). Similarly, \( M_{0x} \) is the mean of features of posts not leading to repost behavior \( y = 0 \). Moreover, \( n_x \) shows the number of posts in the samples except for posts leading to repost behavior \( (y = 1) \), and \( n_0 \) denotes the number of posts in samples which do not result in repost behavior \( (y = 0) \). In addition, \( n \) is the total number of samples examined, \( s_e \) is the standard deviation for values of features of all studied post samples. This correlation coefficient varies between -1 and 1, where 1 shows maximum positive correlation between measured values and user behavior and -1 shows maximum negative correlation between measured values and user behavior. Zero (0) also shows independence of the features from user repost behavior.

\[
r_{pb} = \frac{M_1 - M_{0x}}{s_e} \sqrt{\frac{n_x n_0}{n^2}}
\]

(11)

In the last phase, to evaluate the increasing of information repost behavior in OLC, 5 more methods other than the proposed one, were also employed: (1) latest post method: based on the time a post was created and learners visited the latest \( k \) post, (2) random \( k \) post selection method for a learner, (3) post-post similarity method: \( k \) posts with the highest similarity to the learner’s posts, (4) top \( k \) novel posts: \( k \) post selection method with the highest measured degree of novelty, (5) learner-learner similarity method: \( k \) post selection method with the highest similarity of the posts with the learner. Furthermore, learners were categorized into lurker and posters. Each of these categories is randomly divided into 6 categories. During the 10 day evaluation, each of these 6 categories was received \( k \) post alternatively without advanced notice based on one of these six methods. Equation (12) shows the increasing reposting behavior evaluation metric.

\[
\text{repost rate} = \frac{\text{Number of posts}}{\text{Number of visited Posts}} \times 100
\]

(12)

**Results**

Table (1) shows the evaluation metrics of the proposed repost behavior prediction using logistic regression for lurkers and the posters. The results showed that the proposed method had precisely predicted the learner’s reposting behavior. In order to assess the accuracy of choosing \( k \) effective post, common criteria were used for estimation evaluation.

**Table 1. Repost behavior prediction evaluation**

<table>
<thead>
<tr>
<th></th>
<th>Posters</th>
<th>Lurkers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F measure</td>
<td>Recall</td>
</tr>
<tr>
<td>Proposed Repost Prediction</td>
<td>0.643</td>
<td>0.541</td>
</tr>
</tbody>
</table>
Table (2) presents the results of analysis of Kendall’s correlation coefficient for all posts in the set. Figures 1 and 2 indicate the scatter plot of the estimated novelty and relevance ([0,1]) and observed novelty and relevance (range 1 to 5).

**Table 2. Kendall’s tau correlation coefficient**

<table>
<thead>
<tr>
<th>Relevance (Average of P-P &amp; L-L similarity)</th>
<th>Novelty</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\tau)</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>0.54</td>
</tr>
</tbody>
</table>

![Figure 1. Estimated Relevance (range 0 to 1) Vs. Relevance (questionnaire 1-5)](image)

![Figure 2. Estimated Novelty (range 0 to 1) Vs. Novelty (questionnaire 1-5)](image)

Table 3 presents the results of correlation coefficient between features and repost behaviors. These results are reflective of a positive correlation between these features and repost behavior. The maximum correlation belonged to post-post similarity and topical similarity, whereas the minimum correlation belonged to novelty of posts.

**Table 3. Coefficient of correlation between proposed features and repost behavior**

<table>
<thead>
<tr>
<th>Post-Post Similarity</th>
<th>Novelty</th>
<th>Learner-Learner similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>(r_{pb})</td>
<td>0.63</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.54</td>
</tr>
</tbody>
</table>

According to Table 3, using the proposed method the average of repost rate for each of these two categories, namely, lurkers and posters, was more than those of 5 other methods.

**Table 4. Repost rate for each of methods**

<table>
<thead>
<tr>
<th>Methods</th>
<th>Posters</th>
<th>Lurker</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Repost Rate</td>
<td>Repost Rate</td>
</tr>
<tr>
<td>Random</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>Recent</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>Post-Post Similarity</td>
<td>34</td>
<td>22</td>
</tr>
<tr>
<td>Learner-Learner Similarity</td>
<td>28</td>
<td>17</td>
</tr>
<tr>
<td>Post Novelty</td>
<td>24</td>
<td>20</td>
</tr>
<tr>
<td>Proposed Method</td>
<td>42</td>
<td>34</td>
</tr>
</tbody>
</table>

Also, in order to have a better comparison, a set of one-tailed \(t\)-tests were taken on 10-fold cross validation results. In this experiment, we considered the result of the proposed method as algorithm A2. Hence, if the result of \(t\)-test between the proposed method and another algorithm is a negative value, then the proposed method statistically outperforms the other algorithm. The statistical differences between the experiments obtained with the proposed method, and the other algorithm for the test data sets are listed in Table 5.
Table 5. The result of statistical t-test between the proposed method and other methods

<table>
<thead>
<tr>
<th></th>
<th>Poster</th>
<th>Lurker</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>-3.924</td>
<td>-5.342</td>
</tr>
<tr>
<td>Recent</td>
<td>-79.402</td>
<td>-34.34</td>
</tr>
<tr>
<td>Post-Post Similarity</td>
<td>-29.931</td>
<td>-6.629</td>
</tr>
<tr>
<td>Learner-Learner Similarity</td>
<td>-6.456</td>
<td>-15.686</td>
</tr>
<tr>
<td>Post Novelty</td>
<td>-30.713</td>
<td>-36.554</td>
</tr>
</tbody>
</table>

Discussion

Generally, the topic of increasing the repost of information has been rarely studied in education-related studies despite its great importance in social networks. In this study, in order to answer research question 1, prediction of the effectiveness of the post from the content assessment of candidate posts was emphasized regarding the variety of OLCs and their various capabilities. Features related to the relevance (post-post and learner-learner similarity) and novelty of posts were considered for the prediction of post effectiveness (see Table 1). In research regarding the correlation of relevance and novelty features, a comprehensive assessment was done and the results of the direct effect of these features on learners’ behaviors are shown in Table 3. The results of this study were in line with the results found in various previous research studies such as (Lee & Lee, 2009; Sarma & Panigrahy, 2010; Wang, Duan, Koul, & Sheth, 2014). In addition, the measurement methods of the presented features (relevance and novelty of the post) of the study were compared through the explicit feedbacks of learners in the form of a questionnaire and their accuracies were also assessed (see Table 2).

The role of the proposed method in increasing repost behavior and consequently online participation (the answer to research question 2) was assessed and confirmed for both lurker and poster learners.

Regarding research question 2, both categories of learners had positive repost rates through the selection of effective posts and the display method of this study. The increase in reposts leads to an increase in the participation of learners in online society and sharing of more information. The significance of the improvement through the suggested method compared with other methods is proven in Tables 4 and 5.

Limitation and future research opportunities

The most obvious limitation of this study was the lack of a specific standard in the implementation and design of OLCs. Therefore, the suggested method was designed based on the content of posts to make it applicable in various kinds of OLCs. Other factors such as external events (such as trends and news), context of OLC (such as health forums or technical forums), security concerns and OLC design problems can be effective in learners’ repost behavior. The investigation of these factors is ignored due to the lack of a specific standard and high quality data. Considering at least two kinds of OLCs, future studies can investigate more effective factors in learners’ repost behavior to increase the accuracy of the suggested method. In addition, the display of effective posts under adaptive navigation can be presented for learners based on an educational goal.

Conclusion

OLCs are public or private groups on the internet that address the learning needs of their members by facilitating asynchronous learning. This paper seeks to solve the two fundamental linked challenges being access to the appropriate information and shortage of participation (because of lurking behavior or low participation of some learner) in these communities by selecting $k$ effective posts for learner $v$ among all candidate posts. These posts should be selected based on learner interests, characteristics and knowledge level to motivate the learner to repost it. This reposting behavior facilitates information sharing and increases participation in online communities. For online implementation, the proposed method is suitable to be uses in all kinds of OLCs. Certain features such as post-post similarity, learner-learner similarity, and post novelty are among major parameters taken into account in the selection of $k$ effective posts. The proposed method focuses on the problem of online communities that is lurking learners. Comprehensive evaluations indicate that the proposed method significantly outperformed the existing methods.
Reference


Guest Editorial: Authentic Edutainment with Advanced Technologies

Rustam Shadiev¹, Wu-Yuin Hwang², Gheorghita Ghinea³ and Maiga Chang⁴

¹Nanjing Normal University, Nanjing, Jiangsu Province, China // ²National Central University, Jhongli, Taiwan // ³Brunel University London, Uxbridge, Middlesex, United Kingdom // ⁴Athabasca University, Edmonton, Canada // rustamsh@gmail.com // wyhwang@cc.ncu.edu.tw // george.ghinea@brunel.ac.uk // maigac@Athabascau.ca

Introduction

Advanced information and communication technologies provide great potential for creating new learning environments. The learning process becomes more authentic and educationally entertaining with the help of modern advanced technologies (Shadiev, Hwang, & Huang, 2017). For example, learners experience authentic learning situations with educationally entertaining features, both in the classroom and outside of school, when using advanced technologies. As a result, learning becomes more attractive, effective, and meaningful (Kiernan & Aizawa, 2004; Kramsch, 1993).

Several critical characteristics of an authentic environment were highlighted by Herrington and Oliver (2000) and Newmann and Wehlage (1993). First, this type of environment provides authentic contexts that reflect the way knowledge will be used in real life. That is, learning should take place in a physical environment containing a large number of resources, which preserves the complexity of a real-life setting and reflects the way the knowledge will ultimately be used. Second, it provides authentic activities. Such activities reflect the kinds of activities in which people participate in the real world; they are meaningful and relevant to students and present complex tasks that are completed over a sustained period of time rather than a series of shorter disconnected examples. Third, it creates opportunities for learners to share their learning experiences and to practice with other learners who have various levels of expertise. That is, the students share their experiences and are able to access the experiences of other learners who have various levels of expertise. As a result, the students learn about different perspectives on the topics by considering various points of view and model their skills and performance based on those of experts. Fourth, it offers an authentic learning assessment embedded within the tasks that promotes reflection. The assessment is integrated with learning activities, peer assessment is encouraged, and the learners are assessed based on their outcomes. The learners have the opportunity to compare themselves with other learners who are in various stages of accomplishment and, thus, have the opportunity to improve their own performance and skills.

These characteristics of an authentic environment can be supported by advanced learning technologies (Shadiev, Hwang, & Liu, 2018). Scholars have argued that advanced learning technologies provide a wide range of educational affordances, including the following: pedagogical avails (in situ contextual information, recording, simulation, communication, first-person view), in situ guidance, feedback, distribution and gamification), benefits to educational quality (engagement, efficiency, and presence), and various logistical advantages (hands-free access and free up space) (Bower & Sturman, 2015; Sawaya, 2015). With the aid of advanced learning technologies, authentic edutainment learning environments have been successfully created and used in different fields of knowledge, such as language learning (Huang & Huang, 2015; Lin & Lan, 2015; Liu & Chen, 2015; Shadiev, Huang, Hwang, & Liu, 2018), science education (Looi et al., 2011; Varma, 2014), and mathematics (Carr, 2012; Ross, Morrison, & Lowther, 2010). For example, using mobile technology (e.g., smartphones), students learned basic concepts in the classroom and then applied the newly learned knowledge by solving real-life problems outside of the classroom (Agbatogun, 2014; Lin & Yu, 2016; Lin & Lan, 2015; Liu & Chen, 2015). Current mobile technologies are portable and feature multiple functions (Hwang, Ma, Shadiev, Shih, & Chen, 2016). Such characteristics are useful in supporting the learning process by incorporating many resources from the digital and physical worlds, such as creating multimedia content in an authentic environment, sharing it with classmates and the teacher, studying the content of peers’ work and providing feedback on specific content (Ahn & Lee, 2015; Huang, Yang, Chiang, & Su, 2016; Huang & Huang, 2015; Huang, Shadiev, Sun, Hwang, & Liu, 2017). Furthermore, the learning process can become a healthier and happier experience with the support of wearable devices, such as clothing and accessories that incorporate computers and advanced electronic technologies, e.g., optical head-mounted displays or smartwatches (Bower & Sturman, 2015; Sawaya, 2015). That is, learners are able to not only participate in the learning process but, at the same time, monitor their physical health. Learners are able to make various adjustments throughout learning, such as by becoming more physically active, to achieve a healthier and happier learning experience.

Although many studies have considered the application of advanced learning technologies for learning and instruction, there is not much research that focuses on learning and instruction in authentic edutainment.
environments, where all of the involved components of the environment (the learning environment, the advanced learning system, and/or the adaptive system) support the learning process. Taking this into account, there is a need to propose new approaches, techniques, methods, and processes in the field of authentic edutainment, with the purpose of considering the cognitive and affective aspects of the teaching-learning and decision-making processes.

This special issue aims to present innovative theoretical work and original applications in the field of authentic edutainment. This special issue focuses on original scientific contributions in the form of theoretical and experimental research and case studies that apply new perspectives or theories on the learning processes involved in authentic edutainment. This special issue also includes research on novel technologies that support authentic edutainment.

After a rigorous review process, twelve research papers were selected for inclusion in this special issue. These papers address original scientific contributions in the form of theoretical and experimental research and case studies that apply new perspectives on authentic edutainment.

The first paper, titled Development of a SoLoMo game-based application for supporting local cultural learning in Taiwan, was written by Yen-Ting Lin, Yu-Ming Tseng, Yi-Sheng Lee, Tz-Chi Wang, Shu-I Tsai, and Yun-Jhih Yi. In this paper, the authors aimed to promote students’ local cultural learning. To accomplish this goal, the authors developed a game-based local cultural learning application based on the social, local, and mobile (SoLoMo) principle. The authors investigated the effect of implementing the SoLoMo game-based application on the students’ learning performance in regards to the local culture.

The second paper, User-Oriented EFL Speaking through Application and Exercise: Instant Speech Translation and Shadowing in Authentic Context by Thi-Huyen Nguyen, Wu-Yuin Hwang, Xuan-Lam Pham, and Zhao-Heng Ma, reports on a study on aiding English as a foreign language (EFL) learning with a learning activity that was adapted to students’ interests and the application of the ezTranslate system. The language learners attended language-learning lessons that engaged them in learning while concurrently participating in physical exercise in a real context.

Ting-Ting Wu, Yueh-Min Huang, Chen-Ying Su, Lei Chang, and Yi Chen Lu, the authors of the third paper, titled Application and analysis of a mobile e-book system based on project-based learning in community health nursing practice courses, introduce an e-book system that integrated project-based learning and authentic learning into a community health nursing practice course. The authors explore how the e-book features affected learning interest, motivation, performance and cognitive load.

In the fourth paper, Enhanced Agility of E-Learning Adoption in High Schools written by Gebremariam Mesfin, Gheorghita Ghinea, Tor-Morten Gronli, and Wu-Yuin Hwang, the authors report on a study that investigated the practice of adopting digital media (including a combination of text, images, audio and video) into the school curricula in Ethiopia. The authors surveyed the accessibility of multimedia-rich e-learning resources, the experiences of students and teachers while using multimedia technologies, and their opinions on adopting multimedia in the teaching-learning process.

The paper titled Educational Games to Enhance Museum Visits for Schools by Benoît Bossavit, Alfredo Pina, Isabel Sanchez-Gil, and Aitziber Urtasun focuses on enhancing a visit to the museum through a series of mini-games that shed light on various abstract concepts. The authors selected representative sculptures, designed the corresponding activities, and tested their approach on pupils from primary and secondary schools and students from an educational practice.

Using Exaggerated Feedback in a Virtual Reality Environment to Enhance Behavior Intention of Water-Conservation is the sixth paper and was written by Wei-Che Hsu, Ching-Mei Tseng, and Shih-Chung Kang. The authors utilized an immersive virtual environment technology (IVET) to expose learners to vivid information with personal relevance and immediacy in order to increase their behavioral intention to conserve water. The participants received exaggerated feedback (EF) that intensified the negative consequences of water consumption (direct EF) or environmental damage (ambient EF), which elicited personal affective responses to test the effects of the experimental intervention.

Yi-Lien Yeh, Yu-Ju Lan, Yen-Ting R. Lin report on a study that examined how children collaborate when creating their own stories in a 3D virtual reality (VR) environment and how their collaboration is affected by
The eighth paper of this special issue is titled *A study of using wearable devices for healthy and happy English as a foreign language learning in authentic contexts* and was written by Rustam Shadiev, Wu-Yuin Hwang, and Tzu-Yu Liu. The authors attempted to facilitate English as a foreign language (EFL) learning with a learning activity supported by smartwatches. The activity combined EFL learning with physical exercise, such as walking around the school community. The authors tested the feasibility of facilitating EFL learning while encouraging healthy and happy lifestyle choices by examining the applicability of smartwatches in these contexts.

**Investigating Flipped Classroom and Problem-based Learning in a Programming Module for Computing Conversion Course** is the ninth paper and was written by Adriana Elena Chis, Arghir-Nicolae Moldovan, Lisa Murphy, Pramod Pathak, and Cristina Hava Muntean. The authors investigate the effectiveness of combining the flipped classroom (FC) and problem-based learning (PBL) teaching approaches in a computer programming module delivered as part of a skills conversion course. According to the authors, the combined FC-PBL approach incorporates learning technologies and supports authentic learning by providing an authentic context and multiple perspectives through teamwork and collaboration.

The tenth paper, *Abstract: Game-driven Keyword Auction and Summarization for Academic Reading*, was written by Hercy N. H. Cheng, Calvin C. Y. Liao, and Wan-Chen Chang. In the paper, the authors report on facilitating the reading comprehension skills of graduate students by adopting the reading strategy of summarization under the scaffold of keyword evaluation. In addition, the summarization strategy was transformed into a group-based educational game by incorporating keyword auction mechanisms. The authors aimed to investigate how the summarization approach improves learners’ academic reading skills.

In the eleventh paper, titled *Pedagogical Change in Mathematics Learning: Harnessing the Power of Digital Game-Based Learning*, Siew Pei Hwa focuses on the advantages of multimedia technology and the benefits of digital game-based learning. By using sample lessons from an interactive multimedia courseware called “DigiGEMs,” Siew Pei Hwa emphasizes the use of digital games as a vital tool in mathematics learning. The study sets out to examine if a positive attitude towards the learning of mathematical concepts exists among young learners. It also describes the efficacy of using multimedia and game-based approaches to motivate mathematical learning in young learners.

Finally, the twelfth paper is titled *Gamifying and Mobilising Social Enquiry-based Learning in Authentic Outdoor Environments* and was written by Morris Siu-yung Jong, To Chan, Ming-tak Hue, and Vincent W. L. Tam. The paper describes the development of a mobile application, Gamified Authentic Mobile Enquiry in Society (GAMES), which supports students in conducting authentic outdoor enquiry-based learning in the area of social humanities. The authors report on a quasi-experimental study in which they evaluated the learning effectiveness of GAMES in supporting students’ knowledge construction in comparison to the conventional outdoor enquiry-based learning approach.

**References**


Development of a SoLoMo Game-Based Application for Supporting Local Cultural Learning in Taiwan

Yen-Ting Lin1*, Yu-Ming Tseng2, Yi-Sheng Lee2, Tz-Chi Wang1, Shu-I Tsai1 and Yun-Jih Yi1

1Department of Computer Science, National Pingtung University, Taiwan // 2Department of Computer Science and Information Engineering, National Chung Cheng University, Taiwan // trueblue@livemail.tw // jason8183344@gmail.com // kk196kk196@yahoo.com.tw // scorpion099@kimo.com // MasterYi9202@gmail.com

*Corresponding author

ABSTRACT
Recently, the development of local cultural features and tourism has become important in Taiwan. To support local cultural education, relevant studies have developed outdoor learning approaches and integrated mobile technology to connect real-world and digital-world learning resources. Nevertheless, the above-mentioned developments are usually suitable for a specific learning process and activity. In other words, students may lack the motivation to engage in the same learning process repeatedly after participating in the learning activity. Therefore, to promote students’ local cultural learning, this study developed a game-based local cultural learning application based on the social, local, and mobile (SoLoMo) principle. To investigate the effect of the proposed approach on students’ learning performance with regard to local culture, a quasi-experiment was conducted on a society course at a Taiwanese elementary school. The experimental-group students learned with the proposed approach, while the control-group students learned with the conventional mobile learning approach. The experimental results showed that, compared with the conventional mobile learning approach, the proposed approach significantly improved the students’ learning achievement, learning retention, and learning motivation. Moreover, it was also found that most students showed positive perceptions toward the usage of the proposed application.

Keywords
Local cultural learning, Digital game-based learning, Mobile learning, Social local mobile

Introduction

Recently, the development of local cultural features and tourism has become important in Taiwan. Taiwan’s government has also started to facilitate local cultural education and required mandatory education to develop local culture courses for fostering students’ cultural-historical perceptions and interests.

In the early stages of course development, traditional classroom instruction with textbooks has been a major pedagogy adopted by teachers in the course (Relan & Gillani, 1997). With the advancement and prevalence of information and Internet technology, relevant paper-based resources have been widely digitalized for local cultural education, and students can thus learn the knowledge without the constraints of time and place. Nevertheless, the literature indicates that it is insufficient to support student learning using only digital learning resources because these resources usually lack appropriate learning contexts (Huang, Lin, & Cheng, 2010a; Wei, Lin, & Lin, 2016). Therefore, according to situated learning theory, relevant research has developed outdoor learning approaches to support local cultural education and integrated mobile technology to connect real-world and digital-world learning resources (Hwang & Chang, 2011; Sung, Hwang, & Chang, 2016). Moreover, through the use of mobile devices with RFID or QR codes, a context-aware learning environment can be developed to capture students’ learning status and provide learning guidance or materials (Chiang, Yang, & Hwang, 2014). However, the abovementioned developments are usually suitable for a specific learning process and activity. In other words, students may lack the motivation to engage in the same learning process repeatedly after participating in the learning activity.

Since the development of computer and multimedia technologies, many technology-enhanced learning approaches have been proposed to improve students’ learning motivation (Hung, Hwang, & Huang, 2012; Huang, Yang, Chiang, & Su, 2016; Lu, Chang, Kinshuk, Huang, & Chen, 2014). Among the various technology-enhanced learning approaches, digital game-based learning (DGBL) is an effective educational approach to motivate students to repeatedly rise to the challenge of learning activities (Buckley, Doyle, & Doyle, 2017). The DGBL approach integrates learning strategies, materials, or activities into digital games to develop enjoyable learning environments for teachers and students (Chen & Hwang, 2017). In the past two decades, many educational computer games have been developed to provide a pure digital learning game that did not support real-world interactions (Coller & Scott, 2009; Homer, Plass, Raffaele, Ober, & Ali, 2018; Vasalou, Khaled, Holmes, & Gooch, 2017). In addition, the literature has noted that authentic learning contexts are important to
educate students to reduce knowledge gaps between textbooks and the real world (Wong, Chin, Tan, & Liu, 2010; Zualkernan, 2006). This implies that students possibly lack the ability to apply knowledge to real-world situations while learning without authentic contexts. It is important to integrate the DGBL approach with a mobile learning approach that can connect real-world learning environments with gaming scenarios and motivate students to engage in learning tasks to enhance their learning performance.

As mentioned above, in this study, a mobile, digital, game-based learning application based on social, local, and mobile (SoLoMo) principles is proposed to engage students in real-world learning environments with challenging digital game-based learning tasks to learn local cultural knowledge. In SoLoMo, the word “Local” means the students’ and sites’ geographic information is involved in the learning processes. Local culture represents historical monuments in the student’s hometown. In addition, different from the conventional mobile learning and DGBL approaches, the aim of this study is to develop a mobile digital game using the SoLoMo principle to engage students in game scenarios and motivate them to actively interact with real-world and digital-world learning resources in mobile learning processes. To evaluate the effectiveness of the proposed approach, an experiment was conducted to investigate the following research questions:

- Do the students who learn about local culture with the SoLoMo-based DGBL approach show better learning achievement than those who learn about local culture with the conventional mobile learning approach?
- Do the students who learn about local culture with the SoLoMo-based DGBL approach show better learning motivation than those who learn about local culture with the conventional mobile learning approach?
- Do the students who learn about local culture with the SoLoMo-based DGBL approach show better learning attitude than those who learn about local culture with the conventional mobile learning approach?
- What are the students’ perceptions of the proposed application in terms of ease of use and usefulness?

**Literature review**

**Digital game-based learning**

Digital game-based learning (DGBL) is an instructional approach that integrates learning materials, activities, or strategies into computer or video games. The major aim of DGBL is to effectively motivate students to actively participate in the learning process (Roussou, 2004). In the past two decades, many educational computer games have been developed to provide pure digital games for primary-level to university-level settings across various subject contents (Sung, Hwang, Lin, & Hong, 2017). Van Eck and Dempsey (2002) developed a computer-based simulation game to enhance students’ mathematical skills. Papastergiou (2009) proposed an educational computer game to promote students’ learning motivation and performance for learning computer memory concepts. Beserra, Nussbaum, and Grass (2017) proposed a drill-and-practice video game to support multiple-choice assessment and capture students’ interest and motivation.

To develop an effective educational computer game, researchers noted the following factors: (1) Competition: Competition with other participants or themselves provides motivation to students to engage and achieve a learning task. The competitive element is generally lacking in traditional learning approaches such as a classroom lecture or discussion (Tsai, 2016); (2) Engagement: During the gaming process, students can freely define and modify their gaming strategies according to the specific goals based on the game rules (Hooshyar et al., 2016); (3) Immediate Rewards: In the game, reward settings encourage students to invest time in continuous learning and return to the game to receive more rewards (Beserra et al., 2017); (4) Immediate Reinforcement and Feedback: In the game, students are more likely to learn mistakes since they receive gaming feedback immediately and learn relevant content for challenging the game again.

Furthermore, based on the principle of situated cognition, scholars have stated that educational computer games have positive effects on student learning performance since the learning takes place within a meaningful context (Van Eck, 2006). To date, by using mobile technologies, the DGBL has demonstrated that it is also suitable to support outdoor learning for conducting situated learning activities in the field. With regard to mobile technologies, GPS, RFID, and QR codes are well-known techniques that can be used to acquire students’ geographic information through the use of mobile devices. Therefore, the DGBL with mobile technologies can encourage students to play educational computer games for learning particular knowledge in natural environments. Huizenga, Admiraal, Akkerman, and ten Dam (2009) applied mobile devices with the GPS technique in the field to run a mobile city game to help students learn the historical knowledge of medieval Amsterdam. Chen, Liu, and Hwang (2016) used mobile devices with the QR code technique to implement a context-aware mobile game learning system for assisting students in linking what they have learned in in-field

Social Local Mobile

Social-Local-Mobile (SoLoMo) was coined for business and marketing by John Doerr as the result of the growing popularity of mobile devices and social networks with geolocation techniques (Heinemann & Gaiser, 2014). The SoLoMo principle integrates the user’s social media platforms and geographic location with her/his mobile devices. For education, although previous studies have not formally mention the term SoLoMo, several studies have proposed similar teaching and learning approaches to SoLoMo. For instance, Boticki, Baksa, Seow, and Looi (2015) developed a mobile social location-aware learning system to support self-directed and collaborative learning activities. El-Bishouty, Ogata, Rahman, & Yano (2010) proposed a social knowledge awareness map for a computer-supported ubiquitous learning environment that can detect surrounding environmental objects and then dynamically provide a social knowledge awareness map for peer students according to the detected objects. Huang et al. (2010b) built a mobile learning network based on collaborative services to recommend learning partners with the same interests and specialties to learners and further support learner-oriented mobile learning knowledge networks. Therefore, for education, SoLoMo-based learning applications can provide personalized learning content based on the students’ physical location and connect their learning status with social media platforms. Next, some theoretical perspectives that indicate why this study is beneficial and conducive to developing learning applications with the SoLoMo principle are briefly described with regard to social, local, and mobile aspects.

Social context

Knowledge construction is not confined to an individual; rather, it is a social process between individuals, groups and organizations. From the viewpoint of the social construction of knowledge, Vygotsky posited that the construction of knowledge in a community occurs via the social interactions of its peers (Vygotsky, 1978). This situation occurs frequently in outdoor learning because students have more opportunities to interact with their peers and the environment. Therefore, previous studies have also indicated that outdoor learning is an active interaction process between learners and their environment (Fähræus, 2004; Wilde, Harris, Rogers, & Randell, 2003). Students engaging in outdoor learning can go far beyond the direct supervision and direction of instructors in traditional classrooms. The outdoor learning experience can further encourage students to connect with social networks based on friendship, cooperation, and information exchange in a learning context.

Local context

The explosion of mobile technology has elevated the concept of contextual relevance in the educational field. Students now expect learning services to provide contextually relevant learning resources that inform, edutain or resolve (Lin, 2016). In this study, the SoLoMo-based DGBL approach was adopted as an educational methodology to promote outdoor mobile learning. Not only can it engage students in real-world learning contexts, but it also provides location-aware learning content based on their geographic information. To achieve this aim, this study applied tablets as a learning instrument to instructors and students. This kind of learning device has powerful orientation capabilities (GPS, Bluetooth, NFC, 3G/4G, and WiFi) that can easily capture the physical position of the learning device.

Mobile context

Mobile technologies are indispensable for building the SoLoMo-based DGBL approach and are an essential medium for connecting students, instructors, learning resources, and environments. During the learning process, mobile technologies can help instructors and students to overcome the barriers of time or space to improve the delivery of knowledge (Lin & Lin, 2016). Moreover, through mobile technologies, instructors and students can easily and flexibly conduct learning activities and exchange learning information in the field without the traditional requirement of the learning community being in the same place at the same time (Chee, Yahaya, Ibrahim, & Noor Hassan, 2017). Therefore, this study developed an edutainment application as a mobile media to support mobile technologies that can determine a students’ location to provide learning content that is located nearby and allow students to share learning statuses on a social networking site.
System development

To promote local cultural learning, this study developed a mobile digital game-based learning application based on the SoLoMo principle. The proposed application was implemented by Android programs and integrated mobile techniques (GPS and QR codes), Taiwanese government open data, and social networks. Therefore, instructors and students applied Android-based mobile devices to use the application and conduct outdoor local cultural learning activities. Moreover, in outdoor learning environments, the application captured the geographic information of the students and real-world learning objects and further provided location-awareness learning services to the students. In addition, the students’ learning and gaming achievements could be shared with peers on social networks by using the application.

The proposed application was implemented using the Corona software development kit (SDK), and a database was built using Firebase. Figure 1 shows the architecture of the proposed application. The application was composed of three subsystems: a learning system, a gaming system, and an administrative system.

![Figure 1. The architecture of the learning application](image)

To support the proposed DGBL approach, the gaming system was designed to deliver gaming content to the students. The game style of this application is a kind of tile-matching puzzle game, as shown in Figure 2(a). To improve the students’ learning motivation, this study designed various virtual roles in the game that have different abilities to assist the students in achieving game tasks and further motivating them to collect virtual roles, as shown in Figure 2(b).
Moreover, to promote the students' local cultural learning, the administrative system could be used to detect whether any historic monuments were near the students' location based on government open data in the Taiwan and GPS technique. The administrative system could also guide the students to visit historic monuments in the learning environment and further unlock limited virtual roles and special game tasks for students with regard to a particular historic monument, as shown in Figure 3(a). Furthermore, before starting the special game tasks, the learning system would push relevant learning materials to encourage the students to learn the knowledge about the historic monument, as shown in Figure 3(b). While engaging in the game tasks, as shown in Figure 2(a), the learning system would push the historic monument questions and ask the students to answer the questions, as shown in Figure 3(c). The students could obtain extra power in the game from the gaming system when the questions were answered correctly. The extra power assisted the students in completing the game tasks easily and further collecting limited virtual roles. In addition, the students could share and post personal learning and game achievements with peers on Facebook through the administrative system. Therefore, to collect rare virtual roles
in the game and share individual achievement on Facebook, the students were required to repeatedly visit specific historic monuments and learn relevant knowledge from the historic monuments and the application.

As mentioned above, to store each learning resource and the gaming content, corresponding databases for the proposed application were deployed, including a learning material database, a learning question database, a user profile database, a learning status database, a learning environment database, and a virtual role database. The learning material database is a knowledge collection of historic monuments. The learning question database includes several pieces of information, such as item and answer information. The user profile database stores personal profiles that include students and instructors. The learning status database contains the learning and gaming status of individual students. The learning environment database includes information about historic monuments, such as geographic information. The virtual role database is used to store the information of each role presented in the game.

Experiment

The purpose of the SoLoMo-based DGBL approach is to promote instructors and students to conduct outdoor local cultural learning through the use of the proposed application. This study aims to apply the proposed approach and application to engage the students in achieving each learning target and further assist them in learning relevant knowledge continuously and repeatedly. To evaluate the effect of the proposed approach, the experimental design and settings are described as follows.

Subject

To determine whether the proposed approach truly enhances student learning performance on local cultural learning, a quasi-experiment was conducted on a local cultural learning course at an elementary school in southern Taiwan. The subject of this experiment was conducted on historical monuments in the students’ hometown. The course had a length of 370 minutes, including pre-, post-, and delayed tests. A total of thirty-eight students and an instructor were asked to participate in this experiment. The age of the students ranged between 9 and 10 years of age. All of the students learned the subject first, and they had used mobile devices for 2 years on average. One group of eighteen students served as the control group. The other group of twenty students served as the experimental group. The experimental group was supported by the SoLoMo-based DGBL approach with the proposed application to conduct the course, while the control group was supported by a conventional mobile learning approach without the proposed application. The control group adopted a conventional mobile learning approach rather than a conventional digital game-based learning approach because the purpose of the conventional digital game-based learning approach is to provide a pure digital learning game that does not focus on supporting real-world interactive functions in the game for outdoor learning (Coller & Scott, 2009; Homer et al., 2018; Vasalou et al., 2017). To support the conventional mobile learning approach, a local cultural learning platform was developed, and the learning materials on the learning platform included text descriptions and pictures of the historical monuments. The learning materials and questions for the learning application and learning platform for the two groups were the same.

Research instruments

To evaluate the effect of the proposed approach on student learning performance, various data sources were analyzed, including prior knowledge tests, learning achievement tests, delayed tests, and questionnaire results. The prior knowledge test was designed to assess the students’ knowledge level with regard to the subject of the historical monuments in their hometown before participating in the activity. The learning achievement test was designed to evaluate the students’ learning results after the conclusion of the activity. The delayed test was used to investigate the learning retention of the students after leaving the activity completely. In this study, two teachers were asked to develop the three tests together, and they had each taught the course for more than 6 years. The three tests included 10 multiple-choice test items, and the maximum score of the tests was 100 points. Moreover, the KR-20 reliability of the prior knowledge test, learning achievement test, and delayed test were 0.722, 0.731, and 0.716, respectively. The item discrimination index for most items was greater than 0.35, implying that the items had discrimination validity (Doran, 1980). In addition, three questionnaires were adopted to capture the students’ learning motivation, learning attitude, and acceptance of the proposed application.
With regard to the investigation of the students’ learning motivation, a questionnaire was used from the intrinsic motivation scale of the MSLQ (Motivated Strategies for Learning Questionnaire) developed by Pintrich and De Groot (1990). The learning motivation questionnaire consisted of nine items with a seven-point Likert scale, where “7” represented “strongly agree” and “1” represented “strongly disagree.” The learning motivation questionnaire was used to measure students’ goals and beliefs about the importance and interest of the gaming approach. The pretest and posttest Cronbach’s alpha values of the questionnaire were 0.868 and 0.899, respectively.

Students’ learning attitude was surveyed using a learning attitude questionnaire which has been applied by several studies to measure students’ learning attitudes toward learning activities in various courses (Hwang & Chang, 2011; Lin, Wen, Jou, & Wu, 2014). A total of seven items with a four-point Likert scale were included in the questionnaire, where “4” represented “strongly agree” and “1” represented “strongly disagree.” The pretest and posttest Cronbach’s alpha values of the questionnaire were 0.756 and 0.799, respectively.

The questionnaire for the acceptance of using the proposed application included two scales, “the ease of use of the proposed application” with three items and “the usefulness of the proposed application” with four items (Lin & Lin, 2016; Wu, Hwang, Tsai, Chen, & Huang, 2011; Wu, Hwang, Su, & Huang, 2012). The questionnaire adopted a six-point Likert scale, where “6” represented “strongly agree” and “1” represented “strongly disagree.” The Cronbach’s alpha values for those two scales were 0.785 and 0.706, respectively; the Cronbach’s alpha value of the entire questionnaire was 0.772.

Experimental procedures

Figure 4 shows the experimental process. Students in the experimental group and the control group were asked to take a 20-minute pretest before undergoing the local cultural learning activities. The pretest was conducted to capture the initial learning motivation and learning attitude of the two groups by using the learning motivation questionnaire and learning attitude questionnaire. Moreover, the instructor used 60 minutes to introduce the subject of the historical monuments in their hometown to the students in the classroom. Following that, the students took a 30-minute prior knowledge test, which aimed to evaluate whether the two groups of students had an equivalent basic prior knowledge of the subject learning content.

Before participating in the outdoor learning activity, the students in the control group and experimental group first received a 20-min instruction with regard to the conventional mobile learning approach and the SoLoMo-based DGBL approach in the classroom. Furthermore, the students in the two groups obtained a 30-min lesson with regard to the operation of the mobile learning devices. During the 120-minute outdoor learning activity, the students in the experimental group learned with the proposed application and engaged in the SoLoMo-based DGBL approach. On the other hand, those in the control group learned with the conventional mobile learning approach; that is, the instructor engaged the students in the real-world learning environment for observing and learning about the historical monuments. Simultaneously, the students used the mobile devices to learn using digitalized learning materials from the local cultural learning platform.

The learning environment was an old temple located in southern Taiwan, and there were four learning targets in the temple that needed students to observe and learn relevant knowledge during the learning process. The students in the experimental group and control group visited and observed the temple guided by the learning application and learning platform, respectively. For the students in the experimental group, the learning application unlocked limited virtual roles and four game tasks for the students with regard to the temple. Each game task corresponded to a learning target and involved learning material and two questions with regard to the learning target. The students in the experimental group were asked to engage in the digital game after learning the historical monument knowledge and answer the historical monument questions during the gaming process. For the students in the control group, the learning platform provided the same learning materials and questions provided to the experimental group to the students with regard to the four learning targets. The students in the control group were asked to answer the questions on the learning platform after learning the historical monument knowledge. The students in the two groups conducted the above learning process without constraints. In other words, the students did not need to answer all questions correctly, and they could also learn about other historical monuments.
After participating in all of the learning activities, all students from the two groups received the two questionnaires of the posttest to capture their learning motivation and learning attitude. Additionally, the learning achievement test was used to evaluate the historical monument knowledge they learned in the activities. Furthermore, the students in the experimental group were asked to take the system acceptance questionnaire to survey their perceptions with regard to the ease of use and usefulness of the proposed application. In addition, after three weeks, all students in this experiment were asked to take a delayed test to survey their learning retention with regard to the historical monument. Finally, the entire experimental process was concluded once all procedures were completed.

Results

IBM SPSS was applied to analyze the performance of the students in the experiment, including the results of the prior knowledge test, learning achievement test, delayed test, learning motivation questionnaire, learning attitude questionnaire, and acceptance of the proposed application.

Analyses of prior knowledge, learning achievement, and delayed test

To evaluate the equivalency of the students’ background knowledge with regard to the subject of the historical monuments in their hometown before participating in the learning activities, an independent $t$-test was applied to
analyze the prior knowledge test results between the two groups. Before the analysis, a Shapiro-Wilk test was used to examine the normality of the above data since the participating students were fewer than 50 samples in the experimental group and control group. The value of this test was 0.943 \( (p > .05) \), indicating that the sample satisfied the assumption of normality. Furthermore, Levene’s test for equality of variances was statistically insignificant \( (F = 0.084, p = 0.773 > 0.05) \) and therefore indicates that the group variances could be treated as equal. The t-test result revealed that there was no significant difference between the experimental group and the control group \( (t(36) = -1.221, p = 0.230 > .05) \). The results implied that the students’ prior knowledge with regard to the subject of the historical monuments in their hometown in both groups was statistically equivalent before undergoing the course.

To investigate the effectiveness of the proposed approach for improving the learning achievement of the students regarding the subject of historical monuments in their hometown, one-way independent sample analysis of covariance (ANCOVA) was used to exclude the difference between the prior knowledge of the two groups. To conduct the ANCOVA, the learning achievement and prior knowledge test scores were treated as the dependent variable and covariate, respectively, and the homogeneity of the regression coefficient was tested first. The result confirmed the homogeneity of the regression coefficient \( (F = 0.159, p > .05) \). Table 1 shows the ANCOVA results of the learning achievement for the two groups. The adjusted means and standard deviations were 83.20 and 3.43 for the experimental group and 71.99 and 3.62 for the control group. There was a statistically significant difference between the adjusted means \( (F(1,35) = 4.953, p = 0.033 < 0.05) \). Moreover, the learning achievement of the experimental group was significantly higher than that of the control group. The results reveal that the SoLoMo-based DGBL approach with the proposed application in local cultural learning benefited the students more than the conventional mobile learning approach in local cultural learning with regard to learning achievement.

To further evaluate the learning retention of the students with regard to the subject of the historical monuments in their hometown, the students in both groups were asked to take a delayed test three weeks after completing the experiment. An independent t-test was applied to analyze the scores of the delayed test between the two groups. Before the analysis, a Shapiro-Wilk test was used to examine the normality of the above data since the number of participating students was less than 50 in the experimental group and control group. The value of this test was 0.948 \( (p = 0.079 < .05) \), indicating that the sample satisfied the assumption of normality. Furthermore, Levene’s test for equality of variances was statistically insignificant \( (F = 3.574, p = 0.067 > 0.05) \) and therefore indicates that the group variances could be treated as equal. The t-test result revealed that there was a significant difference between the experimental group and the control group \( (t(36) = 2.177, p = 0.036 < .05) \). The result implies that the proposed approach significantly benefited the students in terms of learning retention.

### Analyses of learning motivation and learning attitude

To analyze whether there were any significant differences between the means of the learning motivation of the two groups after engaging in the entire learning process, ANCOVA was used to exclude the difference between the pretest of the learning motivation of the two groups, and the posttest and pretest scores for learning motivation were treated as the dependent variable and covariate, respectively. The homogeneity of the regression coefficient was not violated \( (F = 0.030, p = 0.863 > 0.05) \). Table 2 shows the ANCOVA results of the learning motivation for the two groups. The adjusted means and standard deviations were 5.78 and 0.11 for the experimental group and 5.39 and 0.11 for the control group. There was a statistically significant difference between the adjusted means \( (F(1,35) = 6.054, p = 0.019 < 0.05) \). The result implies that the proposed approach significantly benefited the students in terms of learning motivation.

### Table 1. The ANCOVA results for the students’ learning achievement

<table>
<thead>
<tr>
<th>Group</th>
<th>Number of students</th>
<th>Mean</th>
<th>SD</th>
<th>Adjusted mean</th>
<th>Adjusted SD</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>20</td>
<td>83.00</td>
<td>9.23</td>
<td>83.20</td>
<td>3.43</td>
<td>4.953</td>
<td>0.033*</td>
</tr>
<tr>
<td>Control</td>
<td>18</td>
<td>72.22</td>
<td>19.57</td>
<td>71.99</td>
<td>3.62</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. *\( p < .05 \).*

### Table 2. The ANCOVA results for the students’ learning motivation

<table>
<thead>
<tr>
<th>Group</th>
<th>Number of students</th>
<th>Mean</th>
<th>SD</th>
<th>Adjusted mean</th>
<th>Adjusted SD</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>20</td>
<td>5.74</td>
<td>0.64</td>
<td>5.78</td>
<td>0.11</td>
<td>6054</td>
<td>0.019*</td>
</tr>
<tr>
<td>Control</td>
<td>18</td>
<td>5.43</td>
<td>0.58</td>
<td>5.39</td>
<td>0.11</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. *\( p < .05 \).*
To evaluate the equivalent of the students’ learning attitude with regard to local cultural learning before participating in the learning activities, an independent t-test was applied to analyze the pretest scores of learning attitude between the two groups. Before the analysis, a Shapiro-Wilk test was used to examine the normality of the above data because the participating students were fewer than 50 in the experimental group and control group. The value of this test was 0.945 ($p = 0.061 > .05$), indicating that the sample satisfied the assumption of normality. Furthermore, Levene’s test for equality of variances was statistically insignificant ($F = 0.380, p = 0.541 > 0.05$) and indicates that the group variances can be treated as equal. The t-test result revealed that there was no significant difference between the experimental group and the control group ($t(36) = 0.042, p = 0.966 > .05$). The result implies that the two groups of students had an equivalent awareness of their learning attitude before entering the course.

In addition, an independent t-test was performed on the rating scores to compare the posttest scores of the learning attitude between the two groups. The values of the Shapiro-Wilk test and Levene’s test were 0.953 ($p = 0.113 > .05$) and 0.594 ($p = 0.446 > 0.05$), indicating that the sample satisfied the assumptions of normality and homogeneity. The analysis result indicated that there was no significant difference in the posttest scores of learning attitude between the two groups ($t(36) = 0.436, p = 0.066 > .05$). The result reveals that the student learning attitude with regard to local cultural learning in both groups was also statistically equivalent after undergoing the learning activities.

### Analysis of system acceptance

The analysis of the acceptance of the proposed application is shown in Table 3, which shows that most students in the experimental group gave positive evaluations. The four questions of the ease of use scale had a mean value exceeding four, indicating that the students evaluated the ease of use of the application positively. It was also observed during the experiment that the students, after being taught once or twice, were familiar with the application operations and learning contexts. For the usefulness scale, all of the questions with a mean value exceeding four represent that a positive evaluation was also provided by the students.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Item</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ease of Use</td>
<td>1. It is easy to operate the mobile device interfaces of the mobile learning application</td>
<td>5.10</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>2. It is easy to read information on the mobile device screens of the learning application</td>
<td>4.95</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>3. The response speed of the learning application is well-matched with the learning progress in the field</td>
<td>5.15</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>4. I think that mobile device operation of the learning application is easy</td>
<td>5.25</td>
<td>0.17</td>
</tr>
<tr>
<td>Usefulness</td>
<td>5. The learning functions provided by the learning application can benefit my learning achievement</td>
<td>5.20</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>6. The operations of the learning application are quite clear and effectively assist me to understand the learning tasks</td>
<td>5.30</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>7. Combining the learning system and the real-world contexts is helpful to learning</td>
<td>5.20</td>
<td>0.20</td>
</tr>
</tbody>
</table>

### Discussion and conclusions

This study developed an Android-based mobile game application with a SoLoMo-based DGBL approach to support outdoor local cultural learning activities. An experiment was conducted to evaluate the effectiveness of the proposed approach. The experimental results showed that, compared with the conventional mobile learning approach, the proposed approach significantly improved student learning achievement, learning retention, and learning motivation. Furthermore, most students in the experimental group agreed with the ease and usefulness of the proposed application in the local culture course.

These findings provide evidence that the proposed approach can benefit students in terms of local cultural learning. From the aspect of learning achievement, the proposed application applied in the proposed approach provided a suitable learning and gaming tool for the students. This result was consistent with the literature because appropriate learning and gaming activities have a significant direct effect on achievement in a DGBL environment (Chen, Wang, & Lin, 2015; Hwang & Wang, 2016). From the view of learning motivation, this...
study designed various virtual roles in the proposed application that could assist the students in collecting roles during the learning process. This design was consistent with past research findings, which noted that effective game design in a digital game-based learning process is an important indicator of student learning motivation (Mathrani, Christian, & Ponder-Sutton, 2016). Furthermore, with regard to learning attitudes, past investigations have indicated that mobile learning approaches and DGBL approaches can improve student learning attitudes (Chao, Chang, Lan, Kinshuk, & Sung, 2016; Chen, Shih, & Ma, 2014; Sáez-López, Miller, Vázquez-Cano, & Domínguez-Garrido, 2015). In this study, the student learning attitudes in both groups were improved, and thus, there was no statistical difference between the two groups. In addition, from the social aspect, the proposed application helped the students to connect with peers that could benefit the students’ learning motivation (El-Bishouty et al., 2010). From the local view, the proposed application engaged the students in an authentic learning environment that enhanced the students’ learning achievement (Chu, Hwang, & Tsai, 2010). With regard to the mobile perspective, the proposed application effectively and efficiently connected the digital-world and real-world learning resources to promote the students’ learning attitude (Hwang, Wu, & Ke, 2011).

To support local cultural learning, several studies have indicated that the SoLoMo-based learning approach can benefit students in terms of their learning performance (Hwang, Chang, Chen, & Chen, 2018; Shih, Chuang, & Hwang, 2010; Sung et al., 2016). Furthermore, a previous study noted that local culture and history subjects are difficult and boring for students since they only memorize the facts about the local culture and history (Zin, Jaafar, & Yue, 2009). Therefore, several studies have suggested that DGBL is an appropriate educational approach to promote students’ local culture and history learning effectiveness (Akkerman, Admiraal, & Huizenga, 2009; Huizenga et al., 2009; Zin et al., 2009). As mentioned above, this study proposed a mobile game to integrate the SoLoMo-based DGBL approach in a local culture course. During the learning and gaming process, the students could only collect limited virtual roles when they were near a particular historical monument and completed special learning tasks. To complete the special learning tasks successfully, the students were required to answer historical monument questions by themselves. These features directed the students to deeply and actively engage in the learning processes. When actively engaged in learning activities and learning processes, students generally show more effective learning achievement and motivation than when learning passively (Hwang et al., 2018; Wood, 2015). Furthermore, with regard to the students’ perceptions of the learning application, although this study did not arrange a formal interview with the students in the experimental group, some feedback was collected from the students during the learning process. In addition to the positive evaluation of the system acceptance from the questionnaire, the students indicated that the user interface and game operation of the learning application were clear and straightforward. In addition, the students stated that they liked to collect the virtual roles in the game since the art design of the virtual roles was attractive to them. The system design mentioned above also benefited the students in actively using the learning application and engaging in the learning process (Wei et al., 2016).

To further discuss the above findings of this study with the literature on SoLoMo-based systems, relevant surveys are presented as follows. Chiang et al. (2014), Hwang, Wu, Chen, and Tu, (2016), and Hwang et al. (2018) proposed augmented reality-based mobile learning systems to support outdoor natural science courses and local culture courses. The mobile learning systems mainly involved learning discussion function (social), location-based guidance function (local), and mobile augmented reality function (mobile) to support outdoor learning. The experimental results indicated that the augmented reality-based mobile learning approach significantly enhanced student learning motivation and learning achievement. The results in the literature are consistent with the results of the present study. Nevertheless, the present study additionally integrated digital game elements into the proposed SoLoMo-based application. This is a major difference compared with the methods reported in the literature regarding motivating students to challenge learning tasks actively and continuously (Buckley et al., 2017; Boctor, 2013). In addition, from a social aspect, the proposed application integrated social network services to facilitate peer connections. From a local view, in addition to location-based learning materials and questions, the proposed application also provided location-based digital game elements to attract students to engage in learning tasks repeatedly. With regard to a mobile perspective, the proposed application applied cloud computing and databases to reduce the device computing load and enhance device mobility.

In sum, the major contribution of this study is to promote local cultural learning for instructors and students. Based on the proposed approach, this study has some limitations and future directions. As the sample size of the experiment was not large, this study needs to conduct experiments with more relevant learning subjects to cover various samples and provide additional evidence. In addition, the short term of the learning session in the experiment is another limitation of this study. To facilitate human-computer interaction in outdoor mobile learning environments, a modern technique, named beacon, will be integrated with the proposed application to.
determine the microposition of the students and further provide intelligent learning services for instructors and students.

Acknowledgements

This study is supported by the Ministry of Science and Technology, Taiwan, R.O.C. under grants MOST 104-2511-S-153-002-MY2 and MOST 106-2511-S-153-003-MY2.

References


User-Oriented EFL Speaking through Application and Exercise: Instant Speech Translation and Shadowing in Authentic Context

Thi-Huyen Nguyen¹, Wu-Yuin Hwang¹*, Xuan-Lam Pham² and Zhao-Heng Ma¹

¹National Central University, Taiwan // ²National Economics University, Vietnam // joynguyen7@gmail.com // wyhwang1206@gmail.com // phamxuanlam@gmail.com // maxma1021@gmail.com

Corresponding author

ABSTRACT

In this study, an application (app) called ezTranslate was developed to aid English as foreign language (EFL) learning by assigning lessons that adapt to students’ interests, in order to motivate them to engage concurrently in physical activity in a real context. ezTranslate was developed by combining state-of-the-art technologies—automatic speech recognition and a cloud translation service, which permit effortless language translation. ezTranslate also offers user-oriented, self-speaking practice opportunities via its speech shadowing function. This app is designed (1) as an effective support tool for EFL speaking in authentic contexts and (2) to engage students in exercise as they walk in their environment. Results suggest that using ezTranslate in familiar situations significantly promotes student cooperation, as evidenced by their increased listening and speaking. Furthermore, there were significant differences in English speaking performance and homework aptitude in participants when using ezTranslate with or without authentic contexts. Indeed, this study’s findings emphasize the importance of authentic support for EFL speaking; namely, students’ interests. The familiarity and fascination of students with real contexts can motivate them to communicate in English better and more frequently. As for the effect of exercise on EFL learning, statistical analysis revealed that the more steps students take, the more they practice speaking English. These results are very promising. Creating an interactive medium in which students not only acquire knowledge but also move their bodies promotes brain and body activity vital to their overall health. In the near future, we plan to extend activity design, executing authentic drama-play to sustain effective, healthy EFL learning.

Keywords

Speech to Speech Translation, Authentic learning context, Healthy learning, Interactive learning environments

Introduction

Today’s rapidly developing mobile technology offers much potential to facilitate foreign language (FL) learning, which can now occur anytime, anywhere, with real contextual support. Thus, learning English as a foreign language (EFL) is more convenient, exceeding the limitations of traditional classrooms. Over the past decade, mobile technology as a learning aid in authentic contexts has been widely investigated. Findings show improved learning among students who applied knowledge to solving the real-life problems typically encountered in familiar and authentic contexts (Golonka, Bowles, Frank, Richardson, & Freynik, 2014) In this case, combining smart devices with several sensors into an educational application (app) that presents an authentic context has great potential to become the future of interactive language learning. According to statistics, speaking English remains challenging for some Asian EFL students; for instance, those in Taiwan, Vietnam, Japan, and China, despite intensive efforts and numerous approaches. From these areas, a few self-confident, self-directed learners rise courageously and actually open their mouths to speak English. However, lack of opportunities to practice English mean they still avoid initiating interactions with English speakers—certainly a disadvantage for their fluency.

Unlike reading and writing which require conscious effort and usually some instruction to learn, children acquire spoken language quickly, easily, and without effort or formal teaching. It happens automatically, whether their parents try to teach them or not. Although parents do not formally teach their children to speak, they do perform an important role by talking to their children. Children imitate sounds from their parents to learn to speak. Eventually, children gain accuracy and fluency in their verbal expression. We can apply the similar strategy for learning English as Foreign Language (EFL) that is creating as much opportunity as possible for students to hear and imitate English. In this study, we developed a mobile translation app called Eztrnanlate which aims to reflect the way children learn their mother language by listening and imitating. This app provides two main functions to help learn English in their surrounding: speech translation and speech shadowing.

The first function, speech translation, is the process by which spoken vocabularies, phrases or simple sentences are instantly translated and spoken aloud in a second language (Wahlster, 2013). We employed this technology to help learners, who have insufficient vocabularies or word forms to express their ideas, translate from their mother tongue to English. With speech translation, learners can experience pronunciation of new words or
sentences immediately, just like the way children listen and learn from their parent’s talk. We assume that EFL learners need to express an idea; they can do it so well in their mother language but not in English. In this case, speech translation can be useful for them to fill this gap, finally speaking their idea in English more accurately. The second function speech shadowing allows learners to repeat (imitate) after the words or sentences spoken by speech translation. By doing so, their English speaking hopefully becomes fluent.

Eztranslate has one advantage due to the mobility of mobile device, which allows students to carry anytime anywhere. However, this advantage often comes with some inconvenience like light weight design, small size of screen and virtual keyboard, hence, sometimes it is not easy to input through such small virtual keyboard. Speech input technology used by Eztranslate can reduce this inconvenience. Eztranslate was designed and implemented for learners to learn English ubiquitously by speaking instead of typing. Inspired from enactivism theory, Interacting with environment is indispensable for developing high level cognition, and meaningful learning is usually drawn from the interaction between learner and environment (Varela, Thompson, & Rosch, 1991), what if we allow students to interact with their real world and bring their own experience into speaking? Is it helpful for students’ speaking performance? In other words, if students are provided with a fully authentic learning environment and use the real world for their speaking, how will their English listening and speaking be affected? Therefore in this study, we designed learning activity with authentic contexts to motivate learners to explore familiar surroundings outside classroom and practice EFL speaking and listening such as going outdoors to make contact with the natural and actual world, experience the weather and facilities of their surroundings and describe them in English. More specifically, students were asked to walk to some of their familiar places and describe them (such as weather, places, people, even, their emotion) in English. During the period of exploring familiar contexts, pedometer is activated to count students’ steps. Hopefully, exploring outside activities help their health development and cultivate the habit of regular physical exercise and movement in learners rather than only sitting in front of computers, then promoting their sense of well-being.

The experiment has been conducted in one university. With a total of 116 first-year university students, the study investigated how ezTranslate could encourage English speaking in familiar contexts, such as outdoor café, campus green garden, sport field, restaurant, dormitory, library, classroom, and how learning activities can engage students in exercise by walking and explore familiar learning contexts, in which they can learn related vocabulary and freely practice daily-life English speaking.

In summary, this study contributes to EFL speaking in three aspects: pedagogy, technology, and psychology:

- First, in pedagogical aspect, the learning activity aims to let students reflect on the natural way children learn their mother language by listening and imitating.
- Second, in technology aspect, we developed a mobile translation app called Eztranslate which provides two main functions to help learn English in their surrounding: speech translation and speech shadowing.
- Third, in psychological aspect, students feel positive and happy when they explore authentic context to practice English speaking. Meanwhile, they like to share their own emotion, sense, and experience which is meaningful and useful for EFL learning.

By doing so, we try to investigate the following research questions:

- Is there any significant difference in students’ speaking performance with and without authentic context?
- What is the relationship among speaking behavior variables (quantity, complexity, and score of students’ speech)
- What is the relationship between physical exercise (number of steps) and speaking behaviors?

Relative literature

Authentic EFL learning with mobile technology

Students are most often engaged in solving real-world problems, and they often express a preference for doing so over listening (Lombardi, 2007). Many studies have shown learning-by-doing as the most effective way to learn. Yet, for a decade, authentic learning has been difficult to implement. However, with the recent and rapid advancement of information and communication technologies, creating authentic learning environments has become easier (Lombardi, 2007). Most people currently own cellphones that offer students an excellent opportunity to learn in the real world. Hwang et al. (2016) suggest that situated authentic contexts contain content-rich learning material, and that, with the aid of mobile devices, students can explore and interact with that material both in and out of the classroom. Besides that, a student is more likely to become engaged in an authentic learning context, especially in a familiar one that they encounter daily and frequently (Golonka et al.,
Technology facilitates foreign language speaking

There have been numerous studies, which showed that technologies contributed directly to improving second language speech and were now being analyzed more closely to help the second language speech production. Typically, technology supports for language speaking skills have two types: Firstly, Computer-Mediated Communication like voice chat or video conferencing, for example, Skype or FaceTime, which requires face-to-face interaction between learners. Video conferencing has become the norm for most tele collaboration projects, tandem learning experiments and social media exchanges. Likewise, synchronous speaking tasks regularly form part of the hybrid or full online language curricula (Blake, 2016).

Secondly, Speech Synthesis Programs like Tell me more, Dragon Naturally Speaking, are useful tools to facilitate self-speaking practice which can strengthen vocabulary and pronunciation (Nomass, 2013). One of the most promising speech technologies which have been used in those programs is automatic speech recognition (ASR). Some recent studies used ASR to facilitate English speaking and found that learners are motivated to say more toward this technology (Lee, 2015; Oberg & Daniels, 2013). ASR is able to capture learners’ voice and then transcribe it into text using speech-to-text technology. Speech-to-text has been proven to be a possible influence on learning outcome for non-native speaker (Shadiev, Hwang, & Huang, 2013; Shadiev, Hwang, Huang, & Liu, 2016) Those combination (of ASR, speech to text, translation, text to speech) has been referred to as “speech translation” or “voice-to-voice translation” and now available in some of the mature services like Google translator.

However, in this study, we developed Eztranslate ourselves with the following reason: Eztranslate was designed for Chinese – English speech translation. At the time of the experiment, not many available apps can do such speech translation except Google Translator and Bing translator, which are commercial software. However, Google Translate or Bing translator is for general translating purpose. We want an app that is specific for English learning. For example, an app can provide score evaluation for students’ speech. Moreover, with Eztranslate, we can easily track and analyze students’ behaviors in detail from our real time server, which is unable to do so with commercial software.

Learning with exercise

Physical activity appeared positively to impact academic achievement and cognitive ability (Conroy, Elavsky, Doerksen, & Maher, 2013; Correa-Burrows, Burrows, Orellana, & Ivanovic, 2014; Fitzsimmons et al., 2014; Hsu, Chen, Su, Huang, & Huang, 2012). For example, Kubesch et al. (2009) observed that 30 minutes of fitness exercise daily improved learning concentration, compared to only 5 minutes of exercise. Kubesch’s study revealed that learning and suitable, ample exercise might be positively related. As for students’ physical and mental health, Käll, Nilsson, and Lindén (2014) found that increasing exercise time did not cause distraction, but released stress and improved the self-efficacy of learners. Many studies have shown that ability and exercise definitely have a strong connection. For example, Raine et al. (2013) observed that exercise greatly improved learners’ memories, while Bass, Brown, Laurson, and Coleman (2013) suggested that schools should not over-emphasize learners’ scores while sacrificing their physical education curriculum. Learners should engage in regular exercise to enhance their physiological and psychological health, thereby improving their reading and understanding abilities for learning. So (2012) found that appropriate exercise can improve blood circulation and stimulate brain activity, possibly affecting academic performance. Scudder et al. (2014) also indicated that exercise improved language learning. In recent years, more and more scholars have suggested that augmented reality-based learning improves learner’s motivation because technology emphasizes their interactions with physical contexts. Hsiao and Rashvand (2011) found that an augmented-reality system not only allows learners to perform side-by-side exercise while learning but also increases their learning pleasure and motivation. Hsiao (2013) also developed an augmented-reality system that enables learners to see the movement of their physical bodies onscreen during simultaneous exercise and learning; this method can boost learners’ interest in learning and improve their time exercising. In addition, Hsiao, Chen, and Huang (2012) has developed an augmented-reality ecology system, with emphasis on the importance of exercise. Results showed that learning and exercising simultaneously not only increased body movement time but also made learning content easier to understand. In the past, however, the relationship between physical activity and learning achievement has not
been deeply investigated. Indeed, most previous research has only roughly measured exercise time, without using sensor technology in wearable or smart devices and has rarely carefully examined the effect on learning.

Methodology

EzTranslate

We developed EzTranslate, a mobile application that runs on an Android platform and is freely available from the Google Play Store. Therefore, we can track students’ learning portfolios like location and interaction from our real-time server and analyze their learning behavior completely. EzTranslate’s goal is to provide a seamless learning experience by using card-based interface (Pham & Chen, 2015; Pham & Chen, 2018) that works on all kinds of mobile devices, such as tablets, smartphones, and smartwatches. It is a combination of the most modest and complex technologies offered by Google, including Google Speech recognition, Google Translate, Text-To-Speech and Google Firebase. These combined features result in immediate speech translation. This ideal combination also delivers a dynamic, seamless learning experience suited to any learning scenario occurring around students.

Speech recognition technology is not a new concept, though the use of mobile devices using speech recognition is increasing day-by-day. Eztranslate uses Google Cloud Speech service (Google, 2015) which has been well tested and evaluated before being formally released for common use. Beside Google’s internal evaluation, some studies also evaluated Google Speech service. For example, A study of (Kudryavtsev, 2015) showed that Google Speech Recognition has better performance under different network conditions compared to Apple Siri. Assefi, Liu, Wittie, and Izurieta (2015) also found that Google voice has the highest percentage of accuracy (nearly 80%) among other four speech services (Nuance AT&T, WIT, IBM Watson). Moreover, during our system implementation phase, 11 Chinese native speakers were asked to evaluate the accuracy of Chinese speech recognition, and they responded that it has very high accuracy (around 95%). Many studies also showed English speech recognition performed very well when English was the first language of speakers (Akhtar, Ali, & Mirza, 2011) However, the accuracy of Chinese speech decreased dramatically for non-native Chinese speakers (70-80%). That was caused by their accent. Similarly, participants in this study who are non-native English speakers felt that “the accuracy of recognizing English sentence speech is low.” This will be a good finding to consider how to implement suitable language learning with concerning regards to the accent of learners’ mother tongue when recognizing their speech.

EzTranslate has four main functions, as follows:

- **Speech Translation**: to translate Chinese speech to English speech or English to Chinese instantly. For example, when a Taiwanese student says, “我愛你,” the system will output “I love you” in sound and onscreen text (Figure 1a). At the top of the screen, an icon allows students to take photos and attach them to translated items.

- **Foreign Speech Shadowing**: to provide a tool for self-regulated speaking. Students can shadow (repeat after) the translated sound “I love you” and get an immediate response as to their accuracy on a 100-point scale (See Figure 1a).

- **History and Map**: Students’ historical speech inputs and outputs, as well as locations where such records occur (GPS data), are saved to students’ local phones and synced real-time into their cloud data servers (see Figure 1b).

- **Homework Submission**: This function allows students to record their voices and submit recordings to their instructors. Students can also replay their submitted speech as many times as they want to (see Figure 1c).

We conducted two experiments in order to answer the three research questions. The first experiment (N = 80) was conducted to investigate the first two research questions and the second one was to study the third research question. This is because the function that records pedometer data was incomplete during Experiment 1. One month later, we conducted a second experiment with another participant group (N = 36) to address the third research question. However, data related to research questions 1 and 2 in the second experiment were subjected to scrutiny. As expected, this part of the data was comparable. However, minor discrepancies led us to certain suggestions in the discussion.
Experiment 1

In Taiwan, 80 first-year university students, aged between 20-21 who enrolled in the course “The Digital Life” for one semester participated in the study. They had smartphones and Internet connections. Initially, all participants were asked to download and install ezTranslate from the Google Play Store. Then, participants were guided in learning how to use ezTranslate and practiced for a few hours. Thereafter, students were given two homework assignments for completion in the following first and second weeks. The homework requirements are (1) Week 1: Assignment 1: Depict in English a favorite place that you know, and (2) Week 2: Assignment 2: Depict in English another favorite place that you know.
For each assignment, students must complete two phases: first, Trial Speak, using speech translation and speech shadowing in the selected context. This phase aims to let students acquire sufficient related vocabulary and sentences and practice their oral skills before entering the second phase. The second phase is Full Speak. Students use the homework submission function (see Figure 1c) to audio record their speech as they describe the favorite place. To guarantee the efficacy of the training, students were asked to join the two phases of the activity at the exact place they intended to describe. By doing so, hopefully, students can easily come up with the ideas from the context to help practice in the first phase, as well as to enrich their speaking assignment in the second phase.

After two weeks, all Trial Speak records and Full Speak audio submissions were collected and analyzed. Post-assignment student questionnaires included two open-ended items to collect their perceptions of ezTranslate. The Figure 2 illustrates the experimental procedure.

Experiment 2

A total of 36 participants, aged 20-27 (excluded students in experiment 1) were arbitrarily recruited from college students who felt interested in using ezTranslate to learn English in familiar context. Half of the students were undergraduate while the other half were graduate students. They were free to use ezTranslate according to their preferred schedules, with no required, pre-selected topics or assignments. For EFL learning, this experiment’s purpose was to discover how our ezTranslate design triggered students’ motivation to walk and explore their familiar contexts. Pedometer data capturing the number of students’ steps were collected during the experiment.

Research variables

The following research variables were collected and studied during the two experiments.

Quantity of Trial Speak: Number of Trial Speak during 1 week. One Trial Speak means the speech student inputs at one time app attempts. Normally, in this practicing phase, student tended to speak one or two short sentence(s) in one attempt.

Complexity of Trial Speak: The definition of complexity of Trial Speak is employed to know how complex students conduct Trial Speak with authentic supported or not. Its example is presented in the Table 1. We measure it via sentence length, (by using a formula of excel software). In English grammar, sentence length refers to the number of words in a sentence. In this study, sentence length was automatically calculated by a formula in excel software. Most readability formulas (methods of measuring or predicting the difficulty level of a text by analyzing sample passages, for example, spell checkers or grammar checkers from Microsoft office products) use the number of words in a sentence to measure its difficulty. Besides that, A recent study of (Zhang, Hwang, Tseng, & Chen, 2018) used the length of T-unit (complexity = Number of words in a sentence/T-unit) to calculate the complexity of the sentences.

<table>
<thead>
<tr>
<th>Table 1. Complexity of a sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presents the degree of complexity of Trial Speak sentences. It counts the quantity of independent words within sentences. For example:</td>
</tr>
<tr>
<td>Today weather is nice.</td>
</tr>
<tr>
<td>Today weather here is nice.</td>
</tr>
<tr>
<td>As weather forecast, it may rain today</td>
</tr>
<tr>
<td>As weather forecast, it may rain and cold today</td>
</tr>
</tbody>
</table>

Average score of Trial Speak: This variable averages all percentage scores. Trial speech is scored automatically by Google Speech service. The score given to trial speech represents the confidence value (between 0-100%) which indicates how confident the Google speech API of the given transcription is correct. For example, a student intends to speak “School,” if the system shows “correct 70%,” it means that Google API has 70% confidence that the student said “School.” If the system shows “Skin – Wrong,” it means that Google API detects that instead of “school,” the student’s pronunciation was more likely “Skin.”

Score of Full Speak: We use 5-point scale criteria for oral speaking (Hughes, 2007) to evaluate Full Speak. This criteria covers 5 dimensions “accent, grammar, fluency, vocabulary, comprehension” which can thoroughly assess the speaking performance of the assignment. Two individual raters, one English Teacher and one
researcher who are both experienced in English with more than 15 years, gave the test scores. There was a high inter-reliability (Spearman’s rho = 0.87) of Full Speak score. The reliability of the evaluation seemed to be fine.

Pedometer: Total steps taken by students while using ezTranslate during the experiment.

GPS: Locations where students used Trial Speak or Full Speak were reported via GPS data and mapped onto a campus map.

Observing student behaviors and data analysis

On our real-time server, we tracked students’ actions via a log file that reported their Trial Speak contents, their locations and their behavior to use the system as well as their Full Speak submissions. Most students completed Trial Speak and Full Speak on campus. GPS data were identified and then automatically plotted onto the campus map. As illustrated in the Figure 3, the most popular locations were classrooms, the food-street, restaurants, dormitories, and outdoor cafés. Numbers in red, blue and yellow circles represent a number of students who completed Trial Speak or Full Speak at these locations. Although we suggested for students to practice English speaking in authentic contexts, however, some students still did not follow this suggestion and practice Trial Speak, or Full Speak in the non-authentic context. Therefore, according to GPS data, Trial Speak logs and Full Speak contents, we categorized students into two groups in term of two phases Trial Speak and Full Speak, respectively as shown in Table 2. For example, in Full Speak activity, we asked students to use ezTranslate to record their speaking related to the topics that they feel interested. According to their content and location analyses, it was found that 44 students did practice in authentic context and while 36 students practiced in non-authentic context.

![Figure 3. Campus locations at which students completed homework assignments 1 and 2](image)

After classification, research variables were analyzed according to the following factors: Quantity, Complexity, and Average Score of Trial Speak between students in group 1A versus students in group 1B, and Average Score of Full Speak of students in group 2A versus students in group 2B.
Table 2. EFL group division with and without authentic context during ezTranslate Trial Speak and Full Speak

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Conditions</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 1 Trial</td>
<td>1A</td>
<td>16 Students who did Trial Speak in authentic context</td>
<td>Talked about basketball in sports stadium or sports field</td>
</tr>
<tr>
<td></td>
<td>1B</td>
<td>64 Students who did Trial Speak in non-authentic contexts</td>
<td>Talked about sport in classroom</td>
</tr>
<tr>
<td>Phase 2 Full</td>
<td>2A</td>
<td>44 Students who did Full Speak in authentic contexts</td>
<td>Talked about food in restaurant</td>
</tr>
<tr>
<td></td>
<td>2B</td>
<td>36 Students who did Full Speak in non-authentic context</td>
<td>Talked about a restaurant in dormitory</td>
</tr>
</tbody>
</table>

Results and discussion

Experiment 1

As shown in Table 3, only two variables, Trial Speak quantity, and Trial Speak complexity showed significant differences between the two groups. Group 1A students spoke much more (M = 49.37, SD = 27.09) than group 1B students (M = 27.09, SD = 39.56). Based on these results, during the 2-week experiment, group 1A performed almost twice as much as group 1B in Trial Speak quantity. Similarly, Trial Speak complexity in group 1A (M = 5.49, SD = 4.8) was much higher than that in group 1B (M = 2.67, SD = 1.3). Complexity of Trial Speak refers to number of words spoken, as mentioned in Table 1.

Table 3. Group description—Trial Speak in ESL students’ authentic and non-authentic contexts

<table>
<thead>
<tr>
<th>Variables</th>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t-value</th>
<th>Sig.(2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantity of Trial Speak</td>
<td>1A</td>
<td>16</td>
<td>49.37</td>
<td>33.24</td>
<td>2.07</td>
<td>.041*</td>
</tr>
<tr>
<td></td>
<td>1B</td>
<td>64</td>
<td>27.09</td>
<td>39.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complexity of Trial Speak</td>
<td>1A</td>
<td>16</td>
<td>5.49</td>
<td>4.8</td>
<td>2.32</td>
<td>.034*</td>
</tr>
<tr>
<td></td>
<td>1B</td>
<td>64</td>
<td>2.67</td>
<td>1.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg. Score of Trial Speak</td>
<td>1A</td>
<td>16</td>
<td>22.39</td>
<td>40.16</td>
<td>1.24</td>
<td>.231</td>
</tr>
<tr>
<td></td>
<td>1B</td>
<td>64</td>
<td>9.26</td>
<td>26.95</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .05.

Results revealed that group 1A students tended to speak complete sentences, while group 1B students were likely to verbalize only vocabulary or incomplete sentences. This revealed that practices in an authentic learning context could promote and engage students to speak more and richer English in each sentence. Particularly, with speech translation, students can be easily stimulated by contexts to internalize English vocabulary or sentences; then they can also “shadow” to practice pronunciation and speaking, often making their English descriptions more vivid and complete. With stimuli from authentic contexts, they can even express English sensory descriptions using sight, sound, smell, touch, and taste to sketch what they see and feel in the authentic context.

A significant difference was found in Full Speak scores between groups 2A and 2B, as shown in Table 4. Group 2A students conducted Full Speak in authentic contexts, while group 2B students did not; hence, Group 2A (M = 3.22, SD = 0.46) outperformed group 2B (M = 2.82, SD = 0.38) (t = −4.185, p = 0.00). However, Trial Speak scores did not differ significantly between groups 1A and 1B.

Table 4. Group description—Full Speak in ESL students’ authentic and non-authentic contexts

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t-value</th>
<th>Sig.(2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. Score of Full Speak</td>
<td>2A</td>
<td>44</td>
<td>3.2</td>
<td>0.45</td>
<td>4.108</td>
</tr>
<tr>
<td></td>
<td>2B</td>
<td>36</td>
<td>2.82</td>
<td>0.38</td>
<td></td>
</tr>
</tbody>
</table>

Note. *** p < .001.

This raises a concern over the notable difference between Full Speak and Trial Speak. In particular, what is the difference in students’ learning behaviors between Trial Speak and Full Speak when they are both used in an authentic context? Further investigation was needed to elaborate this scenario. From our observations, students who used Trial Speak spoke with basic vocabulary, phrases, and sentences, but they spoke longer and in more complete sentences when they used Full Speak. Moreover, from the data log, we found that students normally spent more time (usually the last day before the deadline) seriously accomplishing Full Speak, while Trial Speak was done earlier. Therefore, students were likely to spend more effort on Full Speak, thus explaining the
significant difference between these two speaking scores. To describe an event or context, students were required to look around, feel, think, arrange an idea, and finally express it verbally.

This resembles performing in a drama, but in this case, on the students’ usual, natural “stages.” Thus, in future studies and lessons, purposefully designed, drama-based authentic contexts will likely enhance students’ speaking confidence. In fact, related studies have suggested that drama is a potentially powerful tool for connecting students with the learning process. In addition, by incorporating classroom activities, we have found that activities such as role-playing can provide realistic “experience” of the actual event (Dracup, 2012; Glover, 2014; Kilgour, Reynaud, Northcote, & Shields, 2015; Zhang et al., 2018). By its nature, drama involves students in social contexts in which they are required to think, talk, manipulate concrete materials, and share viewpoints to arrive at decisions (Siks, 1983). Thus, through drama, students can explore both factual knowledge and content concepts while “trying on” social experiences. Some student opinions on authentic context are included below:

“ Able to track the authentic location where I did my practice is very useful for my language learning. Because through authentic learning, I can practice vocabulary, sentences, and whole composition. Eventually, it supports forming a good habit in oral speaking. The most helpful thing is it can conquer the limitation of physical classroom, and I am able to practice English in the interesting authentic contexts.”

“ezTranslate main functions provide ubiquitous translation that can help me to learn English in some different contexts. I think it’s very helpful for learning how to use English in real life.”

“If language learning can be included in authentic context, [it] can speed up the [learning] performance. GPS and map function help me to strengthen connection between language and context. 'Putting English into practice is effective and enjoyable.'

To find connections between Trial Speak and Full Speak, we used the Pearson correlation statistical method. Relationships among variables are presented in Table 5.

<table>
<thead>
<tr>
<th>Table 5. Research variables’ relationships for EFL students using ezTranslate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quantity of Trial Speak</strong></td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>Quantity of Trial Speak</td>
</tr>
<tr>
<td>Complexity Of Trial Speak</td>
</tr>
<tr>
<td>Avg. Score of Trial Speak</td>
</tr>
<tr>
<td>Avg. Score of Full Speak</td>
</tr>
</tbody>
</table>

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

As Table 5 shows, we found two interesting phenomena, explained below:

**Phenomenon 1**

The quantity of *Trial Speak* correlates positively with its average score (0.558, *p* = .000 < .001). We examined log files in order to understand this phenomenon, finding that students normally performed more speaking trials for difficult words that were long and complicated to pronounce. At the beginning, they might have earned low scores, but they repeated these until their scores rose, and, eventually, the average score of their practice increased.

**Phenomenon 2**

The amount of practice time has a significant negative correlation with the average score of *Full Speak* (−0.246, *p* = .027 < .05). This remarkable result also required further investigation. From our observations, we found that students who performed poorly in *Full Speak* tended to practice a lot to improve their ability. However, for them to show marked improvement over such a short time (two weeks) is not easy. Even so, their repeated practice showed that our approach indeed increased their learning motivation. Thus, results show that our approach has
high potential. In the near future, we plan to utilize some mechanisms or activities to help students who performed poorly in Full Speak. Hopefully, their learning can be further enhanced.

**Experiment 2**

The purpose of Experiment 2 was to discover the relationship between physical exercise (in this case, number of steps) and other variables. Therefore, Pearson correlation statistical analysis was applied, and results are presented in Table 6:

<table>
<thead>
<tr>
<th>Quantity of Trial Speak</th>
<th>Complexity Of Trial Speak</th>
<th>Avg. Score of Trial Speak</th>
<th>Pedometer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.104</td>
<td>.508</td>
<td>.347</td>
</tr>
<tr>
<td></td>
<td>p = .544</td>
<td>p = .002</td>
<td>p = .038</td>
</tr>
<tr>
<td>Complexity Of Trial Speak</td>
<td>-0.104</td>
<td>-0.344</td>
<td>-0.157</td>
</tr>
<tr>
<td></td>
<td>p = .544</td>
<td>p = .04</td>
<td>p = .361</td>
</tr>
<tr>
<td>Avg. Score of Trial Speak</td>
<td>.508</td>
<td>-0.344</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>p = .002</td>
<td>p = .04</td>
<td>p = .312</td>
</tr>
<tr>
<td>Pedometer</td>
<td>.347</td>
<td>-0.157</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>p = .038</td>
<td>p = .361</td>
<td>p = .312</td>
</tr>
</tbody>
</table>

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

**Phenomenon 3**

As Table 6 shows, we also found interesting phenomena in this study. In fact, the number of steps significantly correlated with the quantity of Trial Speak (0.347, *p* = .038 < .05), indicating that the more students walk, the more they practice. More specifically, during the experiment, students must find their interesting contexts and then practice speaking English at their chosen venues. With this method, students could practice more while exploring their surroundings. With support from speech translation and speech shadowing, they could easily describe in English what they saw and felt, without the anxiety that usually hinders students’ motivation to speak in class, particularly in Asian countries. In addition, students are motivated to step outside the classroom, meet others, and communicate with them in English. Therefore, it is to say that such a process not only benefits their health but also induces pleasure while they learn English, making it an enjoyable activity as opposed to an intimidating task. Therefore, further investigation of the effects of incorporating enjoyable physical activities into language learning is of great importance. Nature has been promoted as integral to human health and wellbeing, being associated with the capacity to cognitive and emotional wellbeing (Yeh, Stone, Churchill, Brymer, & Davids, 2017). The finding is in line with the previous study that shows that physical activity helps support emotional and mental health, especially green exercise (Green exercise refers to physical exercise undertaken in natural environments). For instance, a study from Focht (2009) compared the effect of brief walks completed in outdoor and laboratory environments on affective responses, enjoyment, and intention to walk for exercise. Participants reported greater pleasant and enjoyment with outdoor walking. Plante et al. (2007) also examined the psychological benefits of exercise environment. Participants were assigned different walking conditions (indoors vs. outdoors), and the result found the experience most enjoyable if outdoors. Several studies also found green exercise reduces anxiety and depressed mood and enhances self-esteem (Pretty et al., 2007; Pretty, Peacock, Sellens, & Griffin, 2005).

Listed below are some opinions from student participants:

"ezTranslate is good for studying English on the go. I can use it every time and everywhere. It provides history and map that I can easily access to review my activity logs and practice some complex words once more. Pedometer and location [function] motivates me to practice English continuously. Burning calories makes learning interesting."

**Phenomenon 4**

Remarkably, the *complexity of Trial Speak* correlates negatively with its *average score* (−0.344 *p* = .04 < .05). In explanation of this phenomenon, students earned higher scores when practicing short or single vocabulary
words, while they earned lower scores when practicing sentences. Of course, pronouncing a vocabulary word is easier than pronouncing a whole sentence. However, in Experiment 1, this correlation showed non-statistical significance (0.036 \( p = .747 > .05 \)), the difference between Trial Speak’s complexity and its average score. The two experiments' differing requirements might explain this. In Experiment 1, students used ezTranslate in one self-selected context, based on an individual student’s interest, to practice and complete Full Speak. Conversely, in Experiment 2, students were free to use ezTranslate as they wished—no homework required—thus possibly leading to different learning attitudes and behavior. More specifically, in Experiment 1, to raise their confidence in Full Speak, students tended to keep practicing a specific number of vocabulary words on their topic. In contrast, students in Experiment 2 tended to “surf,” trying random words or sentences that occurred to them; they also tended to practice speaking arbitrarily. Particularly, the higher the complexity of their Trial Speak, the lower their scores. This situation resulted in negative correlation between complexity and average score of Trial Speak in Experiment 2. These findings emphasize the important role of pedagogical structure in guiding students to practice effectively throughout the study. In other words, pedagogy should always lead technology, first establishing a goal. Likewise, content aligned with learning activities considered suitable tools or functions is a significant instructional strategy for course design. Similarly, as in phenomenon 1 from Experiment 1, another positive correlation was found between quantity and average score of Trial Speak (0.058, \( p = .002 < .05 \)). As shown in table 2, students were likely to practice repeatedly any words or sentences for which they received an early low score. Through repetition, pronunciation scores eventually rose. This progress can explain positive relationships between quantity and average score of Trial Speak.

EFL students’ perceptions of the system and activity design

At the study’s conclusion, we asked students two open-ended questions: (1) Which function of ezTranslate do you like the most? (2) How should we improve ezTranslate? Most students expressed that they liked speech translation and shadowing the most because these features are convenient and helpful for learning English, particularly for low-aptitude students with insufficient vocabulary and grammar knowledge. However, some students also expressed their concerns about (1) speech translation’s accuracy rate and (2) system stability. For example, some students used older model phones that can cause ezTranslate to crash. Other more detailed concerns are included below:

“Speech translation is very functional and handy. The (Google) voice helps me a lot with my listening. Repeating after Google voice speech, shadowing is also a great tool to improve my oral ability. However, sometimes, the accuracy of English sentence is low.”

“ezTranslate allows students to learn everywhere, and the most impressive point is it can provide the (accuracy percentage) rate of my speaking.”

Some students provided suggestions to enhance learning through collaborative activity design:

“I suggest that a multiple player option should be available so that the team players could help one another, wherein anyone who is unfamiliar with a certain word could refer to ezTranslate and their partners and continue their conversation with the other players.”

Discussion and conclusion

In this study, we conducted two experiments. The first examined how ezTranslate could motivate students to perform Full Speak in their familiar, authentic contexts. The second experiment examined how ezTranslate could engage students in walking and learning.

Experiment 1’s results showed that using ezTranslate in a familiar, authentic context motivated students to participate more in a learning activity in which students are more inclined to learn and speak the target language. Students also tended to express more complex, longer sentences. Therefore, information delivered via their speech was also richer in information. Moreover, results also showed that students who completed the final speech in an authentic context outperformed those who did not. Like this study’s trend, research done by (Golonka et al., 2014) also showed that, when students re-enter authentic, familiar environments to learn and practice their skills, background knowledge guides their behavior and helps predict what is to be sought and expected. In such an environment, students are more inclined to learn as they apply new knowledge to solve daily, real-life problems likely to occur in a familiar, natural context. This finding also underlines that ezTranslate, with its speech translation and shadowing, could promote effective authentic learning and, in addition, make learning more interactive and richer in information. In addition, during speech elucidation,
students act as if though on a natural stage, effectively delivering a description of their context; for example, how the food tasted, how cute a girl was, or how nice the weather was. Overall, this study indicates that our mobile learning tool encourages students to participate in a drama-based, authentic learning context to describe their surroundings and provides students with excellent opportunities to develop eloquence in English. Notably, to create effective, authentic learning, students need to use mobile technology according to well-designed pedagogical instruction with clear teacher guidelines.

Experiment 2 revealed that, with ezTranslate, students who took more steps in their authentic context also tended to speak more. This result is optimistic, since creating a healthy learning environment is one of our targets. Exercising is an effective strategy for promoting a healthy body and a good mood. Inspiring students to exercise in any form—walking, running, jumping, playing—contributes to their good health and positive mood, thereby also boosting their learning performance. Recently, the digital game Pokemon Go (Theriault, 2016) became a great example of how technology could engage people to walk and explore their surroundings. Exercise games are thus expected to create many meaningful opportunities for learning (Conlan, 2016; Theriault, 2016). Thus, we should learn to apply such games’ benefits in our education curricula to fuel students’ creativity by promoting advanced language learning; for example, writing English stories and performing through exploration of surroundings. Since the ezTranslate app allows students to snap photos of the real world, they can create and incorporate visual components to make their stories more vivid and relevant. For instance, teachers can ask students to photograph their favorite spot and create a story of why they like it. In this manner, we can induce students’ interest in going outdoors to learn and play.

Limitations and future studies

We acknowledge this study’s major limitation: Students’ actual speaking improvement was not measured since we conducted no pre-test, instead relying on recorded logs and homework performance. While the two experiments were conducted separately, the two types of experiments could be merged. Hence, future studies should conduct a new experiment in which all variables can be examined together. Moreover, pre- and post-tests should assess students’ improvement with the application. Another limitation is the experiments’ short duration. Therefore, performing a longitudinal study to ensure the lasting success of this mobile-based speaking application is necessary because it combines sustainable mobile learning and long-term student motivation. In the future study, we will use these variables such as Trial Speak and Full Speak, Pedometer to analyses their influence on EFL learning performance and their relationship using multivariate statistics. Trial Speak, and Full Speak will the use the same evaluation criteria. Drama-based activity will also be taken into account.

References


Hsu, C.-C., Chen, H.-C., Su, Y.-N., Huang, K.-K., & Huang, Y.-M. (2012). Developing a reading concentration monitoring system by applying an artificial bee colony algorithm to e-books in an intelligent classroom. *Sensors, 12*(10), 14158-14178.


Application and Analysis of a Mobile E-Book System Based on Project-Based Learning in Community Health Nursing Practice Courses

Ting-Ting Wu1*, Yueh-Min Huang2, Chen-Ying Su3, Lei Chang1 and Yi Chen Lu1

1Graduate School of Technological and Vocational Education, National Yunlin University of Science and Technology, Yunlin, Taiwan // 2Department of Engineering Science, National Cheng Kung University, Tainan, Taiwan // 3Nursing Department, National Quemoy University, Kinmen, Taiwan // twu@yuntech.edu.tw // huang@mail.ncku.edu.tw // cysu@nqu.edu.tw // zegxiazhang@gmail.com // m10543006@yuntech.edu.tw

*Corresponding author

ABSTRACT

Nursing practice courses are the most formative part of nursing education. Through practice, nursing students apply, validate, and clarify theories taught in class. Recently, research on the application of mobile devices in nursing practice courses has advanced rapidly. An e-book system that integrates text, audio, and images can improve a learner’s attention, interest, and creativity. Few studies have explored e-book use in education in regional health centers, and studies that have approached this topic have focused only on the data search and retrieval aspects. To address this gap, the current study introduced an e-book system that integrated project-based learning and authentic learning into a community health nursing practice course. Analysis results of a 3-week experiment indicated that the diverse functions, multimedia features, and convenience of the e-book system not only increased learning interest and motivation but also improved learning effectiveness. However, the cognitive load imposed on nursing students using the e-book system also increased, because they were required to create and edit teaching materials for a new health care education method.

Keywords
Community health nursing, Nursing student, E-book, Project-based learning, Authentic learning

Introduction

Both nursing theory and practice are essential to the nursing profession. Nursing students’ capabilities after professional training depended on the effectiveness of nursing education curriculums. Thus, an effective curriculum must be developed to train students before they become professional nurses (DeBack & Mentkowski, 1986; Tseng & Ketefian, 2003). Nursing practice courses are regarded as the capstone of nursing education (Gothler, 1985). Nursing students apply, verify, and clarify theories learned in class through practical experience in the nursing field, thus reducing the potential that they will commit early-career errors when they begin formal practice (Rauen, 1974).

With the emergence of information technology, learning schemes have become more diverse. The prevalence of portable devices has led to new directions in the development of learning models (Christensen & Knezek, 2017; Crompton, Burke, & Gregory, 2017; Hsieh & Tsai, 2017; Karimi, 2016; Kim, Lee, & Rha, 2017). In recent years, applications of portable devices in nursing and practice courses have developed rapidly. Through use of these devices, labor and errors are reduced, and nursing students increase their professional knowledge and skills (Forehand, Miller, & Carter, 2017; Mackay, Anderson, & Harding, 2017; Wu, 2014a). Among these devices, e-books are favorable because they enable integration of text, voice, video, and other media to facilitate the development of social relations among learners and enhance their interest in the learning process. A high-interactivity e-book system can also achieve the effects of edutainment (Sorathia & Servidio, 2012). When used with appropriate teaching strategies, guidance, and planning, e-books serve as a sound tool with which learners can achieve their learning objectives (Muira & Hawes, 2013; Smith et al., 2013; Wu, 2014b). Melrose, Park, and Perry (2013) discovered that with the systematic guidance and real-time assistance of e-books, nurses and other health professionals could implement creative teaching procedures. These new teaching strategies may improve nursing students’ comprehension of course material and advance their practical proficiency. Wu (2014b) developed a multifunction e-book annotation system that nursing students used to analyze and discuss cases among their peers. The study revealed that the introduction of e-books increased learning effectiveness among the nursing students. Nursing students can quickly collect and integrate information from e-books, avoiding the inconveniences of traditional paper-based learning materials. Learning motivation and cooperation among nursing students may also be enhanced through e-book use, which may stimulate the students’ problem-solving and critical thinking abilities. Furthermore, Gueval, Tarnow, and Kumm (2015) introduced a concept-based model into public health courses and incorporated e-books into the traditional curriculum. They reported that most nursing students were satisfied with the immediacy and portability of the e-books. The instant search and review functions of the e-books also led to improvements in the students’ hands-on skills and comprehension, in turn increasing their confidence and motivation.
Nursing courses increasingly emphasize critical thinking, clinical judgment, and problem-solving abilities (Girot, 2000). Teaching strategies affect learners’ message selection, acquisition, and construction, shaping their behavior and thinking in the process (Weinstein & Mayer, 1986). Authentic learning approaches typically focus on complex “real-world” problems and solutions and involve role-playing exercises, problem-based activities, case studies, and participation in virtual practice communities (Lombardi, 2007). These learning environments are inherently multidisciplinary and imitate real-world situations. Authentic learning combines multiple disciplines, perspectives, methods, and communities (Downes, 2007). Students immersed in authentic learning activities can develop the types of “portable skills” that newcomers to any discipline may struggle to acquire on their own (Reeves, Herrington, & Oliver, 2002).

Project-based learning (PBL) was proposed by American educator William Heard Kilpatrick in 1918. PBL has been systematically implemented in medical education to enhance medical school students’ development of critical thinking and problem-solving abilities (Barrows, 1996). Numerous experts have modified the PBL model, which has also been used in nonmedical disciplines (Barrows, 1996; Hou, 2010). PBL involves presenting real-world problems to help students obtain knowledge and skills (Barrows, 1983). This approach can be applied to many disciplines to connect concepts and theories with real-world problems (Barrows & Tamlyn, 1980). Learners develop their metacognitive and critical thinking abilities by solving problems in teams through the collection, analysis, and discussion of information (Krajcik, Blumenfeld, Marx, & Soloway, 1994; Fleming, 2000). In nursing practice courses, group interactions combined with a multifunctional e-book system can increase students’ interest in the learning process, and multimedia presentations can be used to diversify the teaching material of nursing practice courses (Wang, Tan, & Song, 2007; Wu & Chen, 2018). Teaching approaches based on edutainment may enhance students’ learning motivation and active participation (Makarius, 2017; Svane, Aderklot, Fritzdorf, & Hamilton-Jones, 2001).

Most studies on the use of e-books in the medical field have focused on the use of e-books in the operations of large medical institutions and in the conveyance of related concepts (Athilingam, Osorio, Kaplan, Oliver, O’Neachtain, & Rogal, 2016; Farrell, 2016; Jamu, Lowi-Jones, & Mitchell, 2016; Zhao, Freeman, & Li, 2016). E-book use in practical nursing education in regional health institutions, such as district public health centers, has rarely been examined. In addition, with regard to their use in the medical field, e-books have been discussed mostly for their data search and retrieval functions, particularly search and retrieval of disease information, drug information, patient information, and medical guidance (Dee & Stanley, 2015; Ritchie, 2017; Saha, Saha, & Neogy, 2018; Williams & Dittmer, 2009). E-books’ potential uses for integration of teaching theory and practice is seldom mentioned in the literature. To address the gap between theory and practice, traditional teaching methods in nursing practice should be combined with diverse and authentic learning strategies to increase the real-world experiences of nursing students. Edutainment may be used to achieve this purpose. Accordingly, this study introduced an e-book learning system suitable for community health nursing courses. The proposed portable system may be used to diversify learning content and practice activities for nursing practice courses and it offers instantaneous application and support features.

Bradram and Bossen (2003) discovered that when nursing students used e-books, instances in which the devices failed to provide complete information hindered the students’ work and learning processes. Dunphy, Finlay, Lemaire, MacNairn, and Wallace (2011) determined that device malfunction increased users’ frustration in their work processes. Disturbances in accessing information increases mental load, strains working memory, and interferes with long-term memory (Vogel-Walcutt et al., 2011). When the cognitive load required for a task exceeds a learner’s working memory capacity, learning performance is impaired (Young et al., 2014). To determine whether the use of an e-book system adds pressure to nursing practice and whether associated innovations in teaching strategies overburden nursing students during the learning process, this study used a subjective cognitive load scale to assess relevant performance and cognitive load metrics.

In this study, various knowledge areas and skills were integrated in PBL activities to enhance students’ learning motivation and facilitate cooperation and metacognitive development. Under this strategy, learners can acquire knowledge and skills by solving problems and simultaneously cultivate high-level cognitive abilities (Blumenfeld et al., 1991; Solomon, 2003). This study also explored the effect of the proposed e-book system, which integrates PBL and authentic learning, on the learning effectiveness, learning motivation, and cognitive load of nursing students in a community health nursing practice course. Under the strategy implemented in the study, students were encouraged to think critically to address complicated nursing problems, be responsible for their own behavior, and maintain professional and humanitarian attitudes (Watson & Foster, 2003).
E-book system integrating PBL and authentic learning in a traditional community health nursing practice course

The nursing students were tasked with combining both academic and practical skills in a real-world environment after one semester of a community health nursing course. Each nursing student was required to participate in a 3-week practical course concerning community health nursing practices, the content of which included community hygiene. Practical experience of professional nursing can incite nursing students’ interest in the trends of community health nursing. Practical experience also offers an opportunity for students to apply and deepen their understanding of professional information. In the present study, the proposed e-book system was used in both individual health education and community health education sessions. Before the health education sessions, the experimental group of nursing students was required to collate and prepare their materials into interactive e-books. The e-book system enabled nursing students to efficiently create content-rich multimedia interactive teaching materials, and involved lower costs, less time, and fewer privacy issues than those associated with traditional paper methods. Moreover, materials from the Internet were used in the e-books, improving the convenience of data collection and the range of available data. The rich and dynamic content not only drew clients’ attention, but also deepened their impressions of the information, thereby improving the overall administration of health care services. The e-books thus served as edutainment for both the nursing students who developed them and the clients who were their target audience. In the context of this study, “clients” refers to people who received the individual or community health education. The interface of the e-book system developed in this study is represented in Figure 1.

Preparation

Nursing teacher introduced the content, objectives, and plan of the practice program, as well as the purpose and methods of community health nursing. After dividing nursing students into groups, a head nurse of a regional health institution assigned the groups topics regarding individual and community health education. Within their groups, the students were to discuss search methods, labor division, and schedules for the design of the e-book that would serve as the basis for individual and community health education.

Implementation

The nursing students used the e-book system to create health education materials based on personal knowledge and information accessed from the Internet. The software in which students assembled the e-book system was similar to Microsoft PowerPoint; simple drag and drop functions could be used to assemble content. The e-book system also offered numerous predesigned interactive functions, such as drag-and-match, scratch-off, multiple-choice, multimedia play, and picture rotation. Each group used relevant auxiliary functions in the e-book system to create a multimedia presentation and annotate critical parts of the presentation that required further explanation to clients. Students used the Internet to access information related to various health education topics.
Through the collection and analysis of network data as well as communication and discussion within their groups, the nursing students cultivated the abilities to solve problems and synthesize knowledge. Each subgroup in the experimental group collected, organized, and arranged the contents of their e-book according to their assigned topic. The final e-books appeared similar to real books and included a page-turning function and interactive health education content in multimedia presentations. The digitized collation of health education content improved data storage, transmission, management, and use; these improvements also enhanced interpersonal interactions. The nursing teacher assisted students in conducting the relevant activities and guided them in presenting the content. The teacher also observed and assessed student performance.

Presentation

In the implementation phase of this study, students designed and presented individual and community health education materials using vivid interactive multimedia content. The dynamic presentations not only attracted the attention of clients, but also deepened their comprehension of the information, improving the health care service at the basic level. Moreover, the multimedia interactions helped to increase the clients’ interest in health education, strengthened the relationship between clients and nursing students, and boosted students’ confidence in their nursing skills and oral communication abilities. The nursing teacher evaluated the students’ abilities to synthesize knowledge and express themselves orally, and helped them develop skills in problem-solving, communication, and management.

Evaluation

The first goal of PBL evaluation is to assess learning effectiveness. The second goal is the evaluation of the learning process by assessing the problem-solving and interpersonal interactions of nursing students (Shepherd & Cosgriff, 1998). In addition, multidirectional evaluation methods can be used for self-assessment through feedback provided by others (Krajcik, Czerniak, & Berger, 1999). This study used self-reported evaluations from students (25%), a nurse from the internship institution (25%), and the nursing teacher (50%) to assess learning effectiveness.

Reflection

The nursing students reflected on the process of individual and community health education and the corresponding evaluation. Based on their experiences, students considered potential new uses for the system. The nursing teacher assisted the students in improving their methods to obtain better results and increase their understanding.

Research design

Participants

This study enrolled students from two classes at Fourth Nursing School and distributed them randomly into either an experimental group or a control group. Each group comprised 32 students. According to the requirements of the school, the students were further divided into heterogeneous subgroups to complete practical nursing courses. Each subgroup comprised 8–10 students, and groups were led by the same nursing teacher. Nursing students in the experimental group used the e-book system and implemented the teaching strategy of combined PBL and authentic learning. Nursing students in the control group implemented the same teaching strategy, but they used the traditional paper-presentation method. Both groups were required to carry out individual and community health education. To incentivize the nursing students to perform well during the experiment, of the course results were graded.

Experimental procedures

Students participated in the community health nursing practice course in regional public health centers 8 hours a day for 3 weeks. On the first day of the course, the nursing teacher explained the role of public health nurses and the procedure, methods of evaluation, and precautions of practice activities. The nursing teacher separated the
students into groups according to teaching strategy. Time was allotted for students in the experimental group to familiarize themselves with the operation of the e-book system. During the first week, all nursing students were required to attend classes in the regional public health center to learn about the basic functions of grassroots health care and to prepare individual health education materials for home visits the following week. Students in the experimental group were required to create interactive e-books explaining the health education materials (which included information on diabetes, hypertension, chronic kidney disease, and stroke), and students in the control group created paper presentations. During the second week, students were assigned clients for home visits. Students in the experimental group used the e-books to explain basic health information and promote health education to the clients during the home visits, whereas nursing students in the control group used traditional paper-based presentations. The nursing teacher evaluated the performance of students after the home visits. Students reflected on the evaluations they received from the teacher in their preparations of material for community health education sessions (involving topics such as maternal and child health, sexual health, accident injury prevention, chronic disease prevention and treatment, cancer prevention, and elderly care) to be held in the third week. During the third week, each group taught and promoted health education in their assigned community. The experimental group used the multimedia e-book and interactive methods to teach lead community health education, whereas the control group used traditional methods. Figure 2 presents photos taken in the individual and community health sessions.

The nursing teacher evaluated students’ performances in the community health education process. In addition to the assessment from the nursing teacher, the nurse who oversaw the practice course in the regional public health center provided an evaluation score. Before the end of the practice course, the nursing students were required to evaluate their own performance during the practice activities, as well as to rate their learning motivation and cognitive load. Figure 3 presents a flowchart of the experimental procedure.
Assessment tools

Learning effectiveness

This study used three evaluations to determine whether the teaching strategy of the e-book system based on PBL and authentic learning improved learning effectiveness. The nurse from the internship institution who oversaw the practice course, the nursing teacher, and the students evaluated the community health nursing practice process according to seven dimensions. The criteria of these seven dimensions were based on the evaluation of learning effectiveness formulated by the school, and the three evaluation methods were required by the school. The evaluation methods and criteria used in this study have been used in the school for numerous years. This study followed the school’s specifications to familiarize scorers with the evaluation criteria to enhance evaluation accuracy. The seven dimensions are presented in Table 1. Use of multiple assessment sources improved the quality of the evaluation of the learning effectiveness.

Table 1. The seven dimensions of learning effectiveness

<table>
<thead>
<tr>
<th>Performance</th>
<th>I. Basic biological medicine and general care skills</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• I understand what constitutes a healthy community environment and I ensure each client’s safety.</td>
</tr>
<tr>
<td></td>
<td>• I can recognize the causes, symptoms, and risk factors of a client’s illness.</td>
</tr>
<tr>
<td></td>
<td>• I can describe the purpose for a client’s medication as well as how the medication functions and what precautions related to the medication the client should observe.</td>
</tr>
<tr>
<td></td>
<td>• I can recognize when a client’s health status is abnormal (such as abnormal values of three prominent symptoms).</td>
</tr>
<tr>
<td></td>
<td>• I can provide appropriate nursing instructions.</td>
</tr>
<tr>
<td></td>
<td>• I can describe the purposes and methods of various community health care clinics.</td>
</tr>
<tr>
<td></td>
<td>• I can accurately perform various community nursing skills.</td>
</tr>
<tr>
<td></td>
<td>• I can accurately perform the steps of a home visit and complete nursing records.</td>
</tr>
<tr>
<td></td>
<td>• I exhibit excellent performance in my educational hygiene internship.</td>
</tr>
<tr>
<td></td>
<td>• I accurately complete internship assignments on factory hygiene.</td>
</tr>
<tr>
<td></td>
<td>• I can precisely write instructional plans and accomplish group health education tasks.</td>
</tr>
<tr>
<td></td>
<td>II. Care</td>
</tr>
<tr>
<td></td>
<td>• In the nursing process, I apply care skills such as listening and sympathetic engagement.</td>
</tr>
<tr>
<td></td>
<td>• I protect each client’s rights and privacy and actively care about their case.</td>
</tr>
<tr>
<td></td>
<td>• I comprehend and follow ethical rules for nursing.</td>
</tr>
<tr>
<td></td>
<td>• I can participate in ethics-related discussions regarding clinical issues.</td>
</tr>
<tr>
<td></td>
<td>III. Ethics</td>
</tr>
<tr>
<td></td>
<td>• Before performing care procedures, I explain the purpose and process to the client.</td>
</tr>
<tr>
<td></td>
<td>• I respect each client’s self-esteem and value.</td>
</tr>
<tr>
<td></td>
<td>• I protect each client’s personal and illness-related privacy.</td>
</tr>
<tr>
<td></td>
<td>• When executing nursing activities, I can provide the client with fair care rights.</td>
</tr>
<tr>
<td></td>
<td>• When encountering an ethical challenge, I analyze the problem according to various dimensions of ethics.</td>
</tr>
<tr>
<td></td>
<td>IV. Communication and cooperation</td>
</tr>
<tr>
<td></td>
<td>• I can establish a therapeutic interpersonal relationship with a client and the client’s relatives using communication skills.</td>
</tr>
<tr>
<td></td>
<td>• I can actively discuss the care procedure for a client’s illness with the internship instructor or a clinical teacher.</td>
</tr>
<tr>
<td></td>
<td>• I understand the work model of a health care team.</td>
</tr>
<tr>
<td></td>
<td>• I can determine appropriate care service based on a client’s situation and can properly administer care.</td>
</tr>
<tr>
<td></td>
<td>• I can accomplish a task through mutual assistance, cooperation, communication, and negotiation with a team.</td>
</tr>
<tr>
<td></td>
<td>V. Fulfillment of duties</td>
</tr>
<tr>
<td></td>
<td>• I am on time for my professional commitments and do not ask for leaves of absence.</td>
</tr>
<tr>
<td></td>
<td>• I dress appropriately and maintain a professional attitude.</td>
</tr>
<tr>
<td></td>
<td>• I exhibit a diligent, proactive, and serious learning attitude and I seek out learning opportunities.</td>
</tr>
<tr>
<td></td>
<td>• I can manage a community hygiene clinic.</td>
</tr>
<tr>
<td></td>
<td>• I learn actively and I am collaborative. When I discover errors, I report them accurately.</td>
</tr>
<tr>
<td></td>
<td>• I identify with the role and function of a nurse.</td>
</tr>
</tbody>
</table>
VI. Critical and creative thinking
- I can systematically assess a client’s health situation (family and community).
- I can determine the priority of a client’s health problems according their needs (community).
- I can establish and execute personalized care according to the health problems of a client (community).
- I practice effective evaluation according to nursing objectives and measures.
- I use a creative approach to complete community health evaluation reports.
- I use creative thinking to complete group health education activities.

VII. Lifelong learning
- I can proactively provide a client and the client’s relatives with appropriate health education.
- In the psychiatric nursing internship, I proactively collect, read, and use data through multiple channels.
- I demonstrate concern about mental illness issues in society and particular mental disorder cases.
- I determine which aspects of my performance do not meet professional standards, and I adjust the behaviors to further my personal growth.

Learning motivation
The aim of the e-book method was to incite the interest of nursing students, maintain their attention, and stimulate their curiosity and motivation to enhance self-learning efficiency. The effect of the multimedia interactive e-book system on learning motivation was assessed using a modified version of Keller’s (1999) attention, relevance, confidence, and satisfaction (ARCS) learning motivation scale, for which 24 items were rated on a 5-point Likert scale.

Cognitive load
An additional evaluation was conducted to determine whether the introduction of the e-book system engendered additional pressure or burden in the nursing practice process. Students in both groups completed a subjective self-evaluation using a modified version of the cognitive load scale proposed by Kalyuga (2000) and Cerpa et al. (1996), comprising four items rated on a 5-point Likert scale (Table 2). The scale measured two subjective dimensions: mental load and mental effort (Sweller, 1989; Sweller, 1998).

Table 2. Items of the cognitive load scale

<table>
<thead>
<tr>
<th>Mental load</th>
<th>Mental effort</th>
</tr>
</thead>
<tbody>
<tr>
<td>The contents of the tasks were complicated.</td>
<td>I have made an intellectual effort when completing the contents of the tasks.</td>
</tr>
<tr>
<td>The tasks were challenging.</td>
<td>I have given my best to solve the tasks.</td>
</tr>
</tbody>
</table>

Results

Analysis of learning effectiveness
To evaluate learning effectiveness, the independent sample t-test was applied to the scores from the three evaluation types completed based on the practice course results. The effectiveness of integrating PBL and authentic learning through the e-book system into the community health nursing practice course was explored through descriptive statistical analyses.

Table 3. Statistical results of the learning effectiveness analysis

<table>
<thead>
<tr>
<th>Achievement</th>
<th>N</th>
<th>Group</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>p</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-evaluation</td>
<td>32</td>
<td>Control group</td>
<td>86.84</td>
<td>1.886</td>
<td>-4.990</td>
<td>.000*</td>
<td>1.2486</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Experimental group</td>
<td>89.28</td>
<td>2.020</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internship institution evaluation</td>
<td>32</td>
<td>Control group</td>
<td>84.13</td>
<td>1.431</td>
<td>-4.704</td>
<td>.000*</td>
<td>1.1717</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Experimental group</td>
<td>85.78</td>
<td>1.385</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nursing teacher’s evaluation</td>
<td>32</td>
<td>Control group</td>
<td>85.34</td>
<td>1.473</td>
<td>-5.723</td>
<td>.000*</td>
<td>1.4315</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Experimental group</td>
<td>87.56</td>
<td>1.625</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total score of practice course</td>
<td>32</td>
<td>Control group</td>
<td>85.41</td>
<td>1.459</td>
<td>-5.795</td>
<td>.000*</td>
<td>1.4532</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Experimental group</td>
<td>87.55</td>
<td>1.486</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .05.
Figure 4. Average of the three evaluations and total score

As illustrated in Table 3, all p values were less than .05, indicating significant differences between the experimental group and the control group for each of the three evaluation types, as well as for overall score. The average scores from the three evaluations were greater for the experimental group than for the control group. Among the three evaluation types, the evaluation from the internship institution exhibited the lowest scores, and the highest scores were derived from the students’ self-evaluations. The scores closest to the overall average scores were those from the nursing teacher (Figure 4).

Analysis of learning motivation

Learning motivation analysis was conducted by applying the independent sample t-test to the scores of the ARCS scale indicated by the two groups. The reliability of the scale was examined using Cronbach’s α. Based on the ARCS scale scores from the students, the effect of the practice strategies on the four dimensions of learning motivation were determined. The results derived from the reliability analysis of each dimension are presented in Table 4. All values were greater than 0.90, revealing that the learning motivation scale applied in this study exhibited high consistency and reliability.

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Cronbach’s α</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attention</td>
<td>0.963</td>
</tr>
<tr>
<td>Relevance</td>
<td>0.969</td>
</tr>
<tr>
<td>Confidence</td>
<td>0.915</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>0.960</td>
</tr>
</tbody>
</table>

The results obtained from the t-test analysis of the learning motivation scale are presented in Table 5. The value of each item in the four dimensions was less than .05, revealing significant differences between the two groups with regard to learning motivation. A more detailed observation of the average value of each item revealed that for the “relevance” dimension, the difference between the two groups was relatively small, and the “confidence” dimension was associated with the next smallest difference, followed by the difference for the “satisfaction” dimension, and finally, that for the “attention” dimension. These results indicated that the multimedia content and interactivity of the e-book system held students’ attention during the activities and that the teaching strategy of integrating PBL and authentic learning guided the nursing students to achieve the practice objectives. Increased motivation can increase learners’ participation and provide learners with a sense of satisfaction, thereby enhancing self-learning efficiency.

<table>
<thead>
<tr>
<th>Items</th>
<th>Mean</th>
<th>SD</th>
<th>p</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C</td>
<td>E</td>
<td>C</td>
<td>E</td>
</tr>
<tr>
<td>A_1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A_2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A_3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A_4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A_5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A_6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5. Statistical analysis results for ARCS motivation scale scores
Analysis of cognitive load

When a job is assigned to a learner’s cognitive system, it generates cognitive load (Paas & Van Merriënboer, 1993). If the learning process considerably exceeds a learner’s working memory capacity, the learner’s understanding and problem-solving abilities are negatively affected (Sweller, 1989; Sweller, 1998). To understand the cognitive loads that the nursing students experienced during the practice activities, the cognitive load scores provided by the students were analyzed (Table 6).

<table>
<thead>
<tr>
<th>Mental load</th>
<th>N</th>
<th>Group</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>p</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>32</td>
<td>Control group</td>
<td>3.37</td>
<td>.55</td>
<td>-3.325</td>
<td>.0005*</td>
<td>0.8392</td>
</tr>
<tr>
<td>Experimental</td>
<td>32</td>
<td>Experimental</td>
<td>3.53</td>
<td>.57</td>
<td>-3.733</td>
<td>.0005*</td>
<td>0.9354</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mental efforts</th>
<th>N</th>
<th>Group</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>p</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>32</td>
<td>Control group</td>
<td>3.31</td>
<td>.59</td>
<td>-3.643</td>
<td>.0005*</td>
<td>0.9137</td>
</tr>
<tr>
<td>Experimental</td>
<td>32</td>
<td>Experimental</td>
<td>3.46</td>
<td>.56</td>
<td>-3.418</td>
<td>.0005*</td>
<td>0.8746</td>
</tr>
</tbody>
</table>

Statistical analysis (Table 6) revealed that the p values for mental load and mental effort were both less than .05, indicating that the cognitive load and mental effort of the experimental group were greater than those of the control group. Therefore, nursing students using the new e-book method for health education spent more energy and mental resources during the practice activities. In other words, the nursing students that used the e-book system to edit content for individual and community health education experienced higher cognitive loads from this task than the control groups students did from using the traditional method.

Discussion

Practice courses in nursing schools help learners structure abstract concepts, internalize nursing ethics, and apply nursing skills proficiently in real-world situations (Christy, 1980). Nursing practice courses combine theory with practice to facilitate nursing students’ professional development (Clark, Owen, & Tholcken, 2004).

Wireless networks and portable devices have become more advanced and prevalent and have been introduced into nursing practice courses (Huffstutler, Wyatt, & Wright, 2002). Portable devices save labor, reduce errors, and provide instantaneous access to information (Miller et al., 2005). Moreover, the real-time feedback and core knowledge support provided by such devices can contribute to students’ mastery and application of skills (White, Allen, Goodwin, Breckinridge, Dowell, & Garvy, 2005; Wu, 2014a). This study applied an interactive
multimedia e-book system under a teaching strategy of PBL combined with authentic learning in a community health nursing practice course. The rich presentation mechanisms and auxiliary functions of the e-book system enhanced students’ abilities to achieve the practice curriculum objectives. Students were more interested in home visits and the community health education process was more efficient when the experimental method was used.

The results of the learning effectiveness evaluation indicated that the three assessment scores and the overall practice score of the experimental group were higher than those of the control group. Thus, the multifunctional e-book system improved the effectiveness of health education in the context of systematic PBL. Moreover, according to the teacher’s observations, students progressed in self-planning, group cooperation, and practice processes, and they expanded the scope of the health education materials. Among the scores from the three evaluation types, the self-evaluation scores were always the highest, which was consistent with the findings of previous studies (Bouzidi & Jaillot, 2009; Cho, Schunn, & Wilson, 2006; Knowles, 2005). Studies have indicated that evaluation by various third parties is relatively objective, whereas self-evaluation is usually not trustworthy because of personal bias (Bouzidi & Jaillot, 2009; Jonathon, 1986; Novicevic, Buckley, Harvey, & Fung, 2008; Phillips, 2016). Therefore, to assess learning effectiveness, educational institutions should reference a teacher’s or an unaffiliated party’s evaluation. This study used evaluation scores from the practice institution and the nursing teacher as the basis for the overall score of the practice activities. The practice institution scores were the lowest. This evaluation may have been stricter than others because representatives of the practice institution interacted directly with clients. The nursing teacher’s evaluation may be considered the most objective, because the teacher was not connected to clients and could evaluate students objectively.

Learning motivation analysis results indicated significant differences between the perspectives of the students in the two groups. Analysis of the data and the teacher’s observations revealed that nursing students who prepared e-books for individual and community health education were required to use more attention during the practice process than nursing students in the control group. Nursing students in the experimental group required extra time, effort, and resources to refine illustrations and interact with clients. Thus, nursing students in the experimental group scored higher for both mental load and mental effort than those in the control group. These results corresponded with findings in previous studies. Paas (1992) believed that scales for measuring cognitive load and can be used reflect the cognitive cost and to determine the content required for successful learning performance. In addition, Marcus, Cooper, and Sweller (1996) argued that prior knowledge, content organization, and content characteristics in the learning process affect cognitive load. Gerjets and Scheiter (2003) determined that teaching objectives and processing strategies were mediating factors of cognitive load, and that activity planning affected the cognitive load and processing strategies during the learning process. In the present study, nursing students in the control group used traditional paper-based methods to conduct health education sessions and were obliged to carry heavy papers and related materials, in contrast with the portable multimedia presentation and real-time interactive activities of the e-book. The e-book system not only attracted the attention of health education clients but also improved the interactions among nursing students and clients, promoting familiarity and thereby enhancing the effectiveness of the health education service. PBL as a guiding strategy for out-of-classroom teaching may have encouraged the nursing students to increase their initiative, strengthen their problem-solving abilities, and explore solutions through peer cooperation. Furthermore, the majority of nursing students in the experimental group indicated that their listeners appeared excited and interested in the multimedia presentation of the health education content, and thus the nursing students were more willing to participate in promoting health education. The nursing students who used the e-book system felt more professional and confident in the overall practice process, and scored higher than the control group, on average, on every dimension of the ARCS scale.

Conclusion and future work

The introduction of a portable e-book system into community health nursing practice courses can help to reduce the costs, time, and privacy concerns associated with traditional paper methods. Moreover, the Internet can be used in the proposed method to conveniently access and collect a range of information. In the present study, the rich and dynamic content of the e-books not only drew clients’ attention, but also deepened their comprehension of the health education materials, thereby improving health care services at the grassroots level.

The analysis results indicated that nursing students in the experimental group (for which the e-book system was integrated with the combined teaching strategy of PBL and authentic learning) collaborated with each other during the learning process, despite increased cognitive load. They worked diligently, carefully reflected upon their task, and designed their presentations to facilitate positive interactions in individual and community health education sessions. Positive feedback regarding their work also brought them a sense of self-satisfaction. Self-
satisfaction increases learning motivation (Keller, 1999; Small & Gluck, 1994; Lee, 2000). Learning motivation, in turn, increases competence, self-efficacy, positive emotions, expectations, and value, leading to higher performance and learning effectiveness (Cameron & Pierce, 1994; Elliott, & Dweck, 1988; Moos, 2014; Duffy & Azevedo, 2015).

The limitations in this study were that most students in the nursing school were women, and the e-book system was only tested in the community health nursing practice course. Additionally, this study only analyzed learning effectiveness, learning motivation, and cognitive load. In the future, use of the system may be extended to students and teachers of other relevant nursing courses to increase the convenience of their learning environments. More in-depth studies can be conducted on application of the proposed system in specific cases and by specific nursing teachers. The results of this study may serve as a reference for scholars and professionals in relevant fields in their efforts to develop more convenient and effective learning environments.

Acknowledgements

The research reported in this paper has been supported in part by the Ministry of Science and Technology (MOST) in Taiwan under the research project number MOST 105-2628-S-224 -001 -MY3 and MOST 107-2511-H-224 -004 -MY3.

References


Enhanced Agility of E-Learning Adoption in High Schools

Gebremariam Mesfin1,*, Gheorghita Ghinea2, Tor-Morten Grønli3 and Wu-Yuin Hwang4

1Brunel University London, United Kingdom // 2Brunel University, London, United Kingdom // 3Westerdals Oslo ACT, Faculty of Technology Oslo, Norway // 4National Central University, Taiwan // 5Aksum University, Aksum, Ethiopia // gebremariam.assres@brunel.ac.uk, mesfin.assres@gmail.com // george.ghinea@brunel.ac.uk // tmg@westerdals.no // wyhwang1206@gmail.com

*Corresponding author

ABSTRACT

Education is one among the many social services that enjoyed information technology as an enabler. The digital media and e-learning systems have played significant role as a learning content as well as learning platform with the aim to enhance access to education and quality of learning. In this paper, we investigate the practice of adopting digital media (combination of text, images, audio and video) into the school curricula (taking Ethiopia as a case study). We surveyed the accessibility of multimedia-rich e-learning resources; the experiences of students and teachers on using multimedia technologies; and their experiences on adopting multimedia in teaching learning process. Our findings indicated that the experiences of teachers and students on using advanced digital technologies is fair. However, there is only limited access to multimedia rich e-learning resources, and premature practice of adopting the technologies into the teaching learning. The mapping of these results to the existing models of adopting e-learning into the curricula showed that the schools’ status correspond to the initial phases and further recommendations on successful e-learning adoption is provided. In addition, we proposed an agile model to continuously catch the evolving technological innovation and for large scale e-learning adoption into a curricula.

Keywords
E-learning, Multimedia, Agility, Computers in education, Adoption models, Ethiopia, High school, Curriculum, Authentic learning, Edutainment, Advanced learning technologies

Introduction

The proliferation of information technology has made significant contribution to leverage accessible and quality education. For example, computers have been in use assisting the education sector as tools for the learning process; as the entity being studied; and or as instructional content development and delivery platforms (Taylor, 1980; Grundmeyer & Peters, 2016). According to Corbett and Trask (2000), the term e-learning is often used to refer to the development of a computer program or series of programs with explicit aim of replacing the current methods of instruction, often referred to as computer-based instruction; which are traditionally categorized into tutorials, drill and practice, and simulation and games. Today, text, audio, image, video, and their combinations are in use for the development of digital learning content and interactive learning systems; thereby assisting the conventional learning (Horton, 2011). In addition, concepts such as edutainment (Artistikariini, 2017) and authentic learning (Hill, 2016) are shaping the development of advanced e-learning technologies so as to make education enjoyable, and allows students to meaningfully construct concepts and relationships in contexts that involve real-world problems respectively.

Other digital media types (devices and content) can also be incorporated into e-learning. For example, study in (Ghinea & Ademoye, 2010; Dmitrenko & Obrist, 2016) describe additional types of media associated with the sense of smell. Ghinea, Timmerer, and Gulliver (2014) denoted the inclusion of non-conventional media types (e.g., olfaction, haptic, and gustatory) as multi sensorial media, mulsemedia for short. It is also worth mentioning the importance of the mobile technology, virtual reality and augmented reality in education (Merchant et al., 2014).

E-learning is regarded as one educational strategy that can be adopted into a particular subject (Ghinea & Chen, 2006; Panke & Seufert, 2013). Accordingly, its role must be re-evaluated with respect to the pedagogical advantages it provides in components of a course where other strategies are failing. That is, a well-designed and properly adopted e-learning presents materials in an effective way, facilitates communication between teachers and students, and between students themselves; enables collaborative learning activities, easy access to resources, encouraging self-paced learning, and providing online assessment (Anderson, 2002; Bargeron, 2010). Therefore, e-learning can play its part in enhancing the teaching-learning process and enforce a steep learning curve.
In order for e-learning to be effective, it requires planning, academic expertise, and knowledge of information technology. Asan (2003) showed that the extent of introduction of information technology innovation in schools is limited by a number of constraints. This includes, the extent of use of computers beyond the actual computer classes, literacy level of teachers and students on basic skills of using computing devices, knowledge and skill of teachers on the use of computing devices for instructional purposes, availability of expertise for the instructional use of computing devices, and availability of appropriate digital resources. In addition, Abdullah and Ward (2016) regarded that the perceived usability and usefulness of e-learning systems is significant constraints in the design of various technology adoption models.

Several countries tend to incorporate e-learning into their education system. Likewise, in Ethiopia, fragments of computerization initiatives have been carried on at schools with the goal to enhance education quality using e-learning and the awareness on the relevance of information technology in schools by the school community is rising. Thus far, schools have acquired significant number of computing devices through government financing, from donations and the private sector; primarily to supplement the subject of information technology. Mesfin (2004) showed that, although the student to computer ratio is insignificant, schools of all levels at urban areas have introduced computers for office use; a few of them established computer labs for students; and some have Internet access. That is, the computers in schools are mainly used for learning about computers, and for the preparation of teaching materials and examinations.

Significant number of computer literate teachers have also joined the Ethiopian schools through different fragments of teacher development programs (Mesfin, 2004). In addition, there are significant number of technical and vocational education training institutes, colleges and universities training professionals in the area of information technology which would contribute towards providing technology friendly teachers who can adopt e-learning into the curriculum.

Studies such as in (Patel et al., 2017; Mesfin et al., 2017a) highlighted that incorporating e-learning is needed to fill the gaps that exist in the Ethiopian education and enhance the information retention rate. However, to the best of our knowledge, there is a limitation in the existing literature to address the current dynamism in the evolution of digital technology and complement the effectiveness and efficiency of adopting e-learning into the school curricula. Thus, the primary objective of the paper is to explore the existing circumstances in schools (in the Ethiopian context) in terms of students’ and teachers’ access to digital technologies, usage experience and experience of adopting the technologies into the curriculum as variables as described next.

- Access signifies the extent of access to e-learning infrastructure and content both by students and teachers. The down side of this variable could be, for example, lack of access due to unavailability, incompatibility, or insufficient specification (lack of upgrading) for accessing and sharing multimedia rich learning content.
- Usage experience signifies the degree of affinity of teachers and students towards technology which is manifested by acquiring technical skills and attitudes for utilizing digital resources.
- Experience in adopting into curriculum represents the challenges of adopting e-learning technologies into the different topics of a particular course.

In addition, we conducted a review of prevailing e-learning adoption models and mapped the results of assessment of the variables described above - accessibility of e-learning infrastructure and content, the technology usage experience of teachers and students, experience and tendency to enhance the classroom practices using e-learning technologies, particularly in the context of the high schools in Ethiopia.

The findings show that teachers and students demonstrated reasonably fair experience in using digital technologies. However, there is only limited access to multimedia rich e-learning resources, and the practice of adopting them into the teaching learning is premature. The mapping of these results to the e-learning adoption models correspond to the initial phases. In addition, we proposed a phased agile model that would continuously catch the evolving innovations in advanced e-learning technologies and also for successfully applying large scale e-learning adoption. Thus, the agile adoption model benefits high schools so as to effectively integrate advanced e-learning technologies and approaches such as mulsembedia, edutainment, and authentic learning into their curricula.

The paper is organized as follows. The next section presents related work; data analysis, and discussion of results are provided in subsequent sections, respectively; and the last section provides conclusion.
Related work

E-learning technologies

E-learning involves technology, services and content which includes PC based simulation, Web-based training, media streaming, content management systems (CMS), learning management systems (LMS), and hosted digital learning composed of learning objects (Govindasamy, 2001; Henry, 2001; Han & Shin, 2016). Today, LMSs are in use together with learning objects to offer enterprise level design, hosting, and delivery of online courses. For example, Blackboard (see http://uki.blackboard.com/), and Moodle (see https://moodle.org/) are among the LMS that allow authoring multimedia rich learning objects (Yang & Tsung, 2003). Moreover, current e-learning projects such as NEWTON (see http://newtonproject.eu/) aspire for new platforms which would incorporate learning objects which are based on technologies like fabricated labs, gamification, multisensory media, and augmented/virtual reality.

Learning objects are designed for use in multiple training contexts to increase the flexibility of training, and ease of update as described in the SCORM (see http://scorm.com/scorn-explained/) specification (Chiu & Churchill, 2016); content authors store, retrieve and assemble the learning objects into courses (Yang & Tsung, 2003); while the LMS enable to integrate fragmented courses, learner profiling, adhere to standards, testing, and scoring (Yang & Tsung, 2003). In addition, the LMS provides streaming media, access control, calendar, and communication. Such e-learning technologies enable implementation of learner-centered principles in the classroom (McCombs, 2000). Today, concepts such as edutainment (Artistikarini, 2017) and authentic learning (Hill, 2016) are shaping the development of advanced e-learning technologies so as to make education enjoyable, and allows students to meaningfully construct concepts and relationships in contexts that involve real-world problems, respectively.

E-learning in Ethiopian schools

The success of integrating e-learning technologies requires studies in terms of academic success, curriculum, standards, and attitudes of teachers and students (Eib & Mehlinger 1998; Levine & Donitsa-Schmidt, 1997; Mesfin et al., 2017c; Milton, 2003). It is limited by constraints - literacy level of teachers/students, extent of computer-use for instructional purpose, availability of expertise, availability of appropriate digital resources, and perceived usability and usefulness of e-learning systems (Abdullah & Ward, 2016; Asan, 2003; Mesfin et al., 2014; Mesfin et al., 2017b). In this regard, Ethiopia, like many other countries has attempted to adopt e-learning technologies aiming at enhancing education quality in which schools have acquired significant number of computers (some of them established computer networks and Internet) which are mainly used to supplement the subject of “information technology” and prepare teaching materials; and significant number of computer literate (potentially technology friendly) teachers have joined the schools through various teacher/professional development initiatives (Mesfin, 2004). According to Patel et al. (2017), there exist a need for incorporating e-learning in the Ethiopian education to fill the gap by enhancing the learners’ retention rate. However, to the best of our knowledge, the existing literature lacks to address the dynamic evolution in technological innovations and complement the effectiveness of adopting e-learning into the school curricula.

Adoption models

Successful adoption of e-learning technologies in schools require combining different parts of the learning environment which includes a culture of innovation, defining educational goals, the use of technology by teachers and students as a learning tool, working in collaboration for learning; and to become more skillful in choosing own goals, strategies, assessing and monitoring own progress are some of the characteristics that demonstrate successful digital learning adoption (Milton, 2003). As there is a difference between technology acquisition and deployment for intended purpose, the focus needs to be shifted from instructional design to the design of learning environments like teacher development (Kozma, 2000); and various such models indicate that teachers’ awareness of the knowledge society, proficiency in digital resources, imagination for new collaborations, and exposure to complex social dimensions and project-based learning are important for successful adoption (Laferrière, 2009).

The purpose of adopting e-learning into the curricula is to exploit the power of technology to assist learning (Anderson, 2002; Miller, 1997). Thus, the school community (including students and teachers) must accept the fact that computers are part of everyday school life and invest in the design/use of e-learning resources – supply
of resources, technical support, training and reforming the curricula through subject matter re-evaluation (Miller, 1997).

Brunner (1990) and Partee (1996) also pointed out that successful e-learning adoption takes place when technology becomes invisible or transparent and both the teacher and students can concentrate on the content of the course and therefore making it possible for students to use computers in the natural flow of classroom activities. Hence, the impact that e-learning makes in the classroom depends on its availability and upon the ways in which it is used.

The adoption of e-learning into a curriculum depends on the context of the scenario on which it is applied. For example, research in (Alkharang & Ghinea, 2013) identified the lack of management awareness and support, technological barriers, and language barriers as the hindrances of e-learning adoption in higher education.

Several e-learning adoption models have been developed with varying goals and phases. According to Abdullah and Ward (2016), self-efficacy, subjective norm, enjoyment, computer anxiety, and experience of the teacher are among the common factors for the design of technology adoption models; and the perceived ease of use is most affected by self-efficacy while enjoyment is for the perceived usefulness. In this paper, we consider the apple classroom of tomorrow (ACOT), CAMI mathematics, make it happen (MIH) and evolutionary models of e-learning adoption.

**ACOT model.** The ACOT model was developed aiming at documenting how learning and teaching change in a technology-rich classroom (Fisher, Dwyer & Yocam, 1996). The model identified factors which inhibit the changes and support needed to effect fundamental and sustainable change. The ACOT model has entry, adoption, adaptation, appropriation, and invention phases based on the level of introduction of computers in the learning process (Dwyer, Ringstaff, & Sandholtz, 1991).

**CAMI model.** The CAMI model was based on the running of a commercial computerized mathematics school and the implementation of which aim to implement software and new methodology in schools (Miller, 1997). Unlike the ACOT model, CAMI does not focus on technology but on teachers’ skills, hardware, software, and examination results. In this model there is an individualized seat-based learning in the initial phase with the emphasis of improving the retention of knowledge. It has knowledge retention, processing and expansion phases.

**MIH model.** The MIH model is based on an explicit objective set by the school to facilitate collaborative planning and teaching to foster higher-order thinking and it sets the goals of inspiring school leaders to support e-learning adoption; motivate teachers to evaluate and tune their curriculum for inquiry-based learning; and encourage thinking capability, cooperative learning behaviour, and positive attitude towards learning (Zorfass, Remz & Persky, 1991). It has three components - the curriculum, the teacher development and the school facilitation component, which can be implemented over a two years period in three phases.

**Evolutionary model.** The evolutionary model, developed by Miller (1997), is a synthesis of the above mentioned models for a reason that the case study school involved had dominant features of the three models. In this model, successful adoption required teachers’ ownership of necessary technology, skills and role-player capability in the design and evaluation of the curriculum; and the school should make computers part of its planning, train teachers in the method being used in the classroom, and share the goals of e-learning among all the school community. This model has five phases- introduction, entry, intermediate, penultimate, and creation, which does not have a particular time limit and individuals can be at different phase in their e-learning adoption; and transition takes time because it involves people, skills, attitudes, beliefs, organization and finance.

**Ongoing challenges**

Several authors have disclosed the causes for the failures of a number of e-learning initiatives (Saeedikiya, Mooghali, & Setoodeh, 2010; Ssekakubo, Suleman, & Marsden, 2011). For example, majority of e-learning initiatives in developing countries do not fulfil their potential because of low ICT literacy rates among the student community (Ssekakubo, Suleman, & Marsden, 2011); low comfort levels with technology; usability limitation of learning systems; poor marketing strategies; ineffective maintenance strategies and insufficient technical support. These challenges together with the fast change in e-learning technologies has motivated the innovation of many e-learning adoption strategies, activities, and pedagogies (Bidarra & Rusman, 2017; Macleod & Kefallonitis, 2017). However, to the best of our knowledge, the existing literature did not provide a generic e-learning adoption framework that addresses the impact of dynamic technological evolution overtime.
Similarly, Saeedikiya, Mooghali and Setoodeh (2010) mentions the challenges of failure stories of e-learning implementations in terms of high rate of dropout in online courses in universities. The authors suggested a staged approach of adoption (composed of diagnosis, decision making, design, development, delivery, and post-delivery) for implementing in three phases (preparation, operation, and post operation).

Although, nowadays irrelevant for many countries, the above mentioned challenges still persist in many developing countries including Ethiopia. For example, because only insignificant proportion of high schools students in Ethiopia have access to a range of technologies including mobile devices, ICT illiteracy remained an issue.

In general, many of the adoption models focus on mapping the current status of the schools in relation to a presumed full adoption. However, sufficient attention is not given to the fact that attaining complete adoption is an illusion (like that of any automation) and the impact of the ever changing innovation in digital technology on adopting them into the curricula is not given due consideration. Thus, this paper addresses limitations of the existing literature considering the dynamism in the evolution of digital technology and complement the effectiveness and efficiency of adopting into the curricula through a survey of students’ and teachers’ access to digital technologies, usage experience and experience of adopting the technologies into the curriculum.

Methodology and analysis

Methodology

Our approach for the study reported in this paper focuses on the adoption of e-learning in the high schools in Ethiopia. Ethiopia was chosen as a research object because we believe that this is the most convenient (with respect to the first author) to answer our research questions (Marshall, 1996). Thus, following our review of related work, we employed questionnaire as our data collection instrument, and the results were mapped into the requirements of the existing e-learning adoption models presented in the previous section.

The questionnaire was designed for all high school students and teachers (in urban as well as rural areas) and made available online on Google drive. Accordingly, email invitations that state the purpose and containing a link to a Web-based questionnaire were sent out in spring 2017.

It has been pilot tested using responses from ten respondents and validated using the Cronbach’s Alpha (http://www.real-statistics.com/reliability/cronbachs-alpha/) test (the analysis result was 0.77 which is between 0.7 and 0.8). In order to enhance the response rate at a reduced cost and time, we purposively invited more than sixty five (65) potential respondents, who can be easily contacted to respond to the online questionnaires. Thus, we obtained a total of thirty one (31) responses.

The questionnaire is composed of questions (negative and positive phrased) concerning respondents’ access to digital resources, computer skill, e-learning adoption experience, opinions towards approaches for adoption into a curriculum, and an open space for further remarks and the following scoring mechanisms were employed - strongly-disagree, disagree, neutral, agree and strongly-agree.

The demography of respondents was 55% students and 45% teachers. Out of this 32% are information technology teachers; 25% natural science teachers/students; 35% social science teachers/students and 8% teach other subjects. In addition, 22% of all the respondents are personal owners of the computers; 64% use the schools’ computers; and the remaining 14% use computers from other sources.

Overall, our procedure began with review of related work, and data analysis. Next, a mapping of the current status of the schools into the e-learning adoption models was performed. Finally, we employed our study results and provided a description of an agile e-learning adoption model.

Analysis

We performed the analysis using statistical tools including the correlation function to determine the influence of variables (access to and skills on multimedia and computing infrastructure) on the adoption of e-learning. Percentage, mean, and standard deviation are also employed to explore the extent of e-learning adoption,
respondents’ skills, and the dispersion from the mean value, respectively. Thus, the mean responses and standard
deviations corresponding to each variable and to each question is shown in Table 1.

Table 1. Mean and standard deviation of results

<table>
<thead>
<tr>
<th>Variables</th>
<th>Questions#</th>
<th>Total</th>
<th>Teachers</th>
<th>Students</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Access</td>
<td>1</td>
<td>3.839</td>
<td>0.820</td>
<td>3.534</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2.065</td>
<td>0.574</td>
<td>2.076</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4.000</td>
<td>0.516</td>
<td>3.742</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3.129</td>
<td>0.619</td>
<td>2.879</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>3.032</td>
<td>0.948</td>
<td>3.142</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>3.213</td>
<td>0.372</td>
<td>3.075</td>
</tr>
<tr>
<td>Use</td>
<td>6</td>
<td>4.000</td>
<td>0.775</td>
<td>3.777</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>3.807</td>
<td>0.654</td>
<td>3.135</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>3.484</td>
<td>0.724</td>
<td>3.370</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>2.839</td>
<td>0.454</td>
<td>2.468</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>2.484</td>
<td>0.626</td>
<td>2.246</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>3.323</td>
<td>0.295</td>
<td>2.999</td>
</tr>
<tr>
<td>Adoption experience</td>
<td>11</td>
<td>3.000</td>
<td>0.516</td>
<td>2.727</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>1.258</td>
<td>0.576</td>
<td>1.354</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>3.968</td>
<td>0.795</td>
<td>3.777</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>1.968</td>
<td>0.836</td>
<td>1.747</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>2.387</td>
<td>0.882</td>
<td>2.271</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>2.516</td>
<td>0.361</td>
<td>2.375</td>
</tr>
<tr>
<td>Views on approaches</td>
<td>16</td>
<td>4.355</td>
<td>1.050</td>
<td>4.042</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>4.484</td>
<td>0.769</td>
<td>4.011</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>4.355</td>
<td>0.709</td>
<td>3.925</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>4.484</td>
<td>0.626</td>
<td>4.024</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>4.258</td>
<td>0.930</td>
<td>3.928</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>4.387</td>
<td>0.521</td>
<td>3.986</td>
</tr>
</tbody>
</table>

In addition, brief description of the percentages of responses on access, skill and attitude to use, existing
adoption practices, and the respondents’ views on the need for adoption approach is provided next.

Access to digital resources. The first five questions (in the questionnaire) were designed targeting the
respondents’ agreement on the level of access to digital resources. Accordingly, 79% teachers and 47% students
inclined to agree that they have access to digital resources. These results substantiate the work in (Mesfin, 2004)
which pointed out that the important factors for accessibility are somehow available - pupil-computer ratio,
connectivity and availability of educational software. This is because high schools are used to receive computers
for the information technology course and a few computers are connected as local area networks and to the
Internet; and satellite based network for broadcasting digital multimedia. In most cases, the computers are used
to have document editors, spreadsheets, and drawing tools; sometimes encyclopedia, tutorial, drill games, and
test applications for subjects such as mathematics, physics and language specific tools can be found.

Usage experience. The usage experience factor, which focuses on the skills and attitudes of the respondents in
digital technologies, was also analyzed based on the responses on a set of five questions in the questionnaire.
Thus, the mean responses to the five questions 76% and 79% agreement for having perceived skills and attitudes
by the students and teachers, respectively.

Adoption practices. Another five questions were also dedicated to investigate the experience of students and
teachers on using digital resources for learning purposes. The mean responses of agreements for the five
questions were 0% and 21% by students and teachers, respectively. These results reiterate the fact that furnishing
schools with digital resources and advanced learning technologies does not necessarily imply successful
adoption (Mesfin, 2004).

Views on adoption approach. Adopting computers in education has positive contribution towards student’s
learning outcome (Ghinea, & Chen, 2008; Hadley, & Sheingold, 1993). But, we concentrated on the challenges
of dynamic technological change (e.g., advent of advanced technologies) and the holistic myth for adopting e-
learning resources into the curricula; and the respondents’ views on adoption approaches they foresee were
collected from the responses of other five questions in the questionnaire. Accordingly, the levels of agreements on the questions targeting these concerns were 100% and 93% by students and teachers, respectively.

*Correlation.* The correlation between respondents’ access to digital resources and adoption experience; and computer usage experience and adoption experience into the learning was analyzed. Accordingly, the correlation coefficients of the mean responses of agreements of both respondent types for access and usage experience were -0.17 and 0.09, respectively.

Finally, the independent samples *t*-test result for the two groupings (access versus integration, and usage versus integration) is shown in Table 2. In Table 2, *p*-value of the mean differences in scores of the responses corresponding to both groupings is zero which shows that the difference in mean scores between the groups is statistically significant.

**Table 2. Independent samples *t*-test result**

<table>
<thead>
<tr>
<th>Levene’s Test for Equality of Variances</th>
<th><em>t</em>-test for Equality of Means</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>F</em></td>
<td>Sig.</td>
</tr>
<tr>
<td>Access vs Integration</td>
<td>179.515</td>
</tr>
<tr>
<td>Usage vs Integration</td>
<td>68.150</td>
</tr>
</tbody>
</table>

Accordingly, based on this statistical result, a mapping of the existing status of the schools in Ethiopia into the e-learning adoption models is described in the next section. In addition, an e-learning adoption model (in the context of advanced e-learning technology) is proposed, and further discussion on the overall result is provided in subsequent sections.

**Mapping into adoption models**

Based on the analysis result provided above, we performed a mapping of the schools’ current status in the process of e-learning adoption. The mapping is achieved with reference to the adoption models described in the previous section - ACOT, CAMI mathematics, MIH, and Evolutionary. Although adoption models such as in (Abdullah & Ward, 2016) are more recent, these models are more relevant and hence selected because they attempt to address specific issues about schools.

Accordingly, we employed these models to map and reflect on the scenario of adopting digital learning technologies into the schools with respect to the responses from the online questionnaire (see Table 3). Table 3 depicts each phases of the models shaded with green, yellow and orange background representing the extent of the mapping-fitness of the schools status as fit, moderate, and unfit, respectively.

The goal of the ACOT model is to provide understanding on how technology would affect teaching and learning; hence requires technology rich environment (Fisher, Dwyer & Yocam, 1996). It assumes that teachers will be more likely taking the next step in integrating digital learning technologies into their teaching when their comfort with technology increases (Ertmer & Ottenbreit-Leftwich, 2010). Arguably, the schools in Ethiopia in general are not well equipped with digital learning technologies. However, the high schools have acquired relatively enough as discussed above which substantiate to correspond more to the entry phase (and somehow to the adoption phase) on the ACOT model. In addition, the outlier responses (standard deviation) could perhaps be because teachers use computers to prepare their lecture notes; but it is not sufficient condition to fit in the adaptation phase.

The transition between phases in the CAMI model is assumed to be triggered by the change in the skills of teachers and availability of technology (Miller, 1997). As the knowledge retention phase does not require sophisticated technology, supplementing lectures using simple drill software is a satisfactory condition to fit in. However, such drill software (in addition to document edition) is only intermittently employed which makes it somewhat fit into the retention phase.

In the MIH model, the need for deliberate integration of digital learning requires reorganization of the school to support integration, a team of organized teachers study, experiment and tune the curriculum (Zorfass, Remz &
The deliberate integration of technology into the school environment might correspond to the SchoolNet initiative in Ethiopia. However, the lack of curricular re-evaluation as to how certain learning objectives can exclusively be supported by the technology, makes it unfit even to phase one of the model.

### Table 3. A mapping of Ethiopian high schools into adoption models

<table>
<thead>
<tr>
<th>Phase</th>
<th>ACOT</th>
<th>Evolutionary</th>
<th>CAMI</th>
<th>MIH</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>Entry</strong> - Technology is installed; Staff is unfamiliar; Use traditional methods</td>
<td><strong>Introduction</strong> - Computers and connectivity introduced; teachers understand technology; training begins; some teachers attempt to use in teaching</td>
<td><strong>Knowledge retention</strong> - limited skill; drills used to retain recently taught subjects; small networked system</td>
<td><strong>Phase one</strong> - Curriculum evaluated; topics selected for integration; vision shared; inquiry-based learning software; innovative teachers select for piloting</td>
</tr>
<tr>
<td>2</td>
<td><strong>Adoption</strong> - High computer access; drill supplements lecture; traditional methods</td>
<td><strong>Entry</strong> - Drill software supports teaching; teachers drop fear of technology; small change in class layout</td>
<td><strong>Knowledge processing</strong> - sufficient software for data processing; skillful teachers; teachers can plan and prepare</td>
<td><strong>Phase two</strong> - The same team (phase 1) continues implementing in the first year</td>
</tr>
<tr>
<td>3</td>
<td><strong>Adaptation</strong> - Integrated into classroom; new cognitive outcome; Traditional lecture</td>
<td><strong>Intermediate</strong> - experiment and use for learning; teachers investigate teaching strategy; look for new systems and class rearrangement</td>
<td><strong>Knowledge expansion</strong> - sufficient hardware and software; better skills; organizational planning skills of teachers</td>
<td><strong>Phase three</strong> - The old team continues to plan, design and evaluate the curriculum; adds more teams; mentors the new teams</td>
</tr>
<tr>
<td>4</td>
<td>** Appropriation** - Teachers mastery of computers; collaboration and interaction; self-paced; multi-modal; change of timetable for team work</td>
<td><strong>Penultimate</strong> - Team teaching develops; school timetable changes; better computer skills; teacher’s role changes to facilitator and looks for better technology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td><strong>Invention</strong> - Intensive access to computers; high interaction; learning by doing</td>
<td><strong>Creation</strong> - Teachers gain the capacity to decide which technology fits their practices</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.** Green, yellow and red representing the extent of fitness of the schools status as fit, somehow fit, and unfit, respectively.

The evolutionary model, on the other hand, requires teachers to have high access to computers and the schools’ plans for gradual integration digital learning into the curricula. In this regard, the statistical results described above show that the curricula lack prescriptive guidelines for adopting digital learning which makes the schools status to somehow fit on the introduction and entry phases.

Overall, the existing models only map the e-learning technologies available during the mapping for the high school context in Ethiopia and no advanced learning technologies which may be introduced in the future are considered. Thus, we propose new e-learning adoption model as described next.

### Agile adoption approach

In our analysis, respondents have opted for an adoption approach that embraces the dynamically changing technological innovations and also recognized the need for gradual progression (as opposed to big bang) in the process of adoption. Thus, we propose an agile adoption model that insists prescriptive guidelines through explicit curricular re-evaluation; which takes the access to relevant e-learning resources and usage opportunities for certain experience of state-of-the-art technologies that teachers and students have as a driver.

Agile e-learning is not a new concept. For example, Tesar and Sieber (2010) suggested an e-learning development approach that focuses on personalized learning, usability of learning utilities, learner centered design, and flexible course concepts based on the manifesto of the agile software development. However, to the
best of our knowledge, there is no e-learning adoption model that considers the need for dynamically responding to the ever changing technological innovations. Accordingly, we coined the continuous technology adoption approach depicted in Figure 1. In addition, we adopted agile principles for integrating the dynamically changing e-learning technologies into the curricula (see Table 4).

The approach in Figure 1 takes into account the fact that the digital technology continues to evolve throughout the lifetime of the curriculum in question. Thus, an evolving technology adoption team corresponding to the curriculum must always be available watching new technological innovations, listening to the needs of teachers and students, thereby continuously adopting e-learning into the curriculum.

In this model, the task of e-learning adoption is participatory process that takes students’ new culture and abilities, and trained teachers (learning content and platform) and results into an updated prescriptive curricular document for adopting technology into learning objective(s) of the subject matter(s) under review. The updated curricula serves as a binding rule to be enforced particularly regarding the adoption of technology into the learning practice.

The task of curricular re-evaluation is performed iteratively similar to the iteration in the agile software development (Paulk, 2002). The activities performed in an iteration include need identification for the adoption of technology into certain learning objectives, selection of appropriate off-the-shelf digital content for the learning objective, development of adoption guideline, pilot testing, and documenting the new curriculum which is potentially approvable. Thus, each iteration produces an approvable prescriptive curriculum in one semester duration.

In this model, we adopted the agile principles for software development (Paulk, 2002). Table 4 describes the adopted principles; which are the basis for the e-learning technologies adoption model. This model is different from the model suggested in (Tesar, & Sieber, 2010) in that the latter is based on the agile manifesto. In this model, the curriculum re-evaluation team is analogous to software development team while the primary customers are teachers and students.

When the adoption model is adopted, the tasks of curriculum re-evaluation, technology identification, and developing scenarios continuous as one mainstream business of the school through the curriculum’s lifetime. Once certain learning objective is chosen for piloting, it enters into a fast track iteration of re-evaluation, technology selection, scenario design and pilot testing within one semester duration. The success of an agile iteration is determined by an approval of the pilot for full scale implementation as of the upcoming semester.

The agile model of e-learning adoption is composed of a set of principles, a process, and a set of phases. The set of principles shown in Table 4 leverage the agile e-learning adoption process. The set of activities constituting the process are initiating e-learning adoption for a curriculum; planning for adopting e-learning in one or more selected small chunk(s) of learning objective(s) to be implemented over a semester; perform weekly reviews aiming at removing backstopping; and conduct retrospective meeting to evaluate the process and demonstrate contribution of the adoption in learning. Once approved, e-learning will be formal integral part of the learning
objective for the upcoming semesters; subsequently other sets of learning objectives are selected for testing another adoption during the next semester; and so on.

Like the adoption models presented in the previous sections, the agile model has a set of phases. Since each adoption of e-learning in the chunks of learning objectives is well tested; and satisfies the characteristics that demonstrate successful e-learning adoption (Milton, 2003). That is, students and teachers involved in the selected learning objectives will be able to acquire a culture of innovation, define educational goals, use technology as learning tool, work in collaboration for learning, create own delivery strategies, and provide assessment and monitoring own progress. Thus, the phases in the agile adoption model represent the percentage of learning objectives re-evaluated for e-learning adoption (as per the requirements of the agile process) out of the total number of learning objectives constituting the curriculum in question; and the phases could be, for example, zero, twenty, fourth, sixty, eighty, and hundred percent completion.

Table 4. Agile principles for adopting e-learning into a curriculum

<table>
<thead>
<tr>
<th>Principles</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Priority</td>
<td>The highest priority is satisfying the needs of teachers and students through early and continuous delivery of re-evaluated curriculum, learning technologies and scenarios.</td>
</tr>
<tr>
<td>Change</td>
<td>Welcome curriculum, technology and scenario change requests at all times.</td>
</tr>
<tr>
<td>Delivery</td>
<td>Deliver re-evaluated curriculum, technology and scenario every semester.</td>
</tr>
<tr>
<td>Lifetime</td>
<td>Pedagogical experts and curriculum evaluators must work together throughout the lifetime of the curriculum.</td>
</tr>
<tr>
<td>Teaming</td>
<td>Build an adoption team of motivated individuals by providing convenience, support, and autonomy.</td>
</tr>
<tr>
<td>Conversation</td>
<td>The most efficient and effective method of conveying information to and within an e-learning adoption team is face to face conversation.</td>
</tr>
<tr>
<td>Measure</td>
<td>Pilot tested adoption scenario is the primary measure of progress.</td>
</tr>
<tr>
<td>Sustainability</td>
<td>The adoption team, teachers, students and principals must ensure sustainable technology adoption by maintaining a constant adoption pace indefinitely.</td>
</tr>
<tr>
<td>Attention</td>
<td>Continuous attention to subject matter expertise and new technology enhance agility.</td>
</tr>
<tr>
<td>Maximization</td>
<td>Identify and maximize the amount of learning objectives for which adoption scenarios that can be developed in the future.</td>
</tr>
<tr>
<td>Self-organization</td>
<td>The best scenarios emerge from self-organizing technology adoption teams.</td>
</tr>
<tr>
<td>Reflection</td>
<td>At regular intervals, adoption teams reflects on how to become more effective, and adjust behavior.</td>
</tr>
</tbody>
</table>

Overall, the agile adoption model takes advanced learning technologies and concepts which may be introduced in the future into account. Thus, it benefits the high schools in Ethiopia so as to effectively integrate advanced e-learning technologies and approaches such as multimedia, edutainment, and authentic learning into their curricula.

Results

Here, the contexts presented in the above sections are further discussed. Accordingly, the results corresponding to the accessibility of computing devices, usage experience, adoption experience, correlations, the mapping, and implications for the agile model are presented next.

Accessibility. The statistical data in the above section showed that respondents do not have severe shortage of access to digital resources. In fact, the mean responses on access for reading conventional documents (e.g., hypertext and pdf) is higher than for playing multimedia-rich content (e.g., videos); and their agreement on the regular hardware/software upgrade to support up to date multimedia content is generally lower than the mean. The respondents’ agreements are perhaps because high schools are used to get computers from initiatives like supplementing practical sessions of the subject of information technology, empowering vocational information technology courses, and the SchoolNet initiative. Despite the increase in distribution, however, the culture of continuous refurbishment is not yet attained. On the contrary, respondents have easy access to multimedia content. Overall, the results signify that there is no acute shortage of access to e-learning resources which would hinder adoption into the curricula. However, issues like unavailability of state of the art digital resources, incompatibility, or insufficient specification (lack of upgrading) for accessing and sharing multimedia rich learning content; which all require a responsive mitigation approach.
**Usage experience.** The statistical data showed that respondents have the desired skills and attitudes towards document editing, data processing, searching for digital content, exploring online learning, and collaborative working with email; and the mean responses to all of the questions relating to usage experience specify agreement. This indicates that the degree of affinity of teachers and students towards technology (manifested by acquiring the necessary technical skills and attitudes for utilizing digital resources) is satisfactorily attained.

**Experience in adoption.** The experience of teachers and students on the use of digital resources in learning is generally limited to document editing and Internet searching. That is, teachers usually perform document editing and Internet searching for preparing lecture notes; and students for their assignments. Furthermore, the statistical result reveals that the experience of teachers in encouraging their students to utilize computers in their learning is low. The implication is that, in many cases, students utilize digital resources either as part of the subject of information technology or for social purposes. In general, the experience of teachers and students in adopting digital resources in learning as compared to the access and usage experience is low. This, perhaps, might be because of lack of prescriptive curriculum for e-learning adoption.

**Correlation.** Our statistical result showed low correlation between the accessibility of computing resources, and adoption experience, and skill of computer usage and adoption experience. This signifies that further effort is required to re-evaluate the curricula for successful e-learning adoption.

**Mapping.** Accordingly, our discussion of results (above) and the mapping of the current situation in the schools into the adoption models showed that having access to (and skill on) computers did not correspond to the adoption experience. Thus, the schools are found to resemble at most to the initial phases of the models implying that the development of new e-learning adoption model is paramount importance.

**Agile model.** As described in the section on agile adoption, the phases in the agile adoption model depend on the percentage of learning objectives (in a specific curriculum) which are re-evaluated for the adoption of e-learning. Thus, the status of the schools with respect to the model requires a study for an aggregate result of each learning objectives of the curricula corresponding to the schools.

Overall, the study reported in this paper showed significant insight into an advanced approach for successful agile and continuous e-learning adoption into the school curricula. However, the limitations of this research is that the impact of the agile adoption model is not yet tested in the context of specific curricular re-evaluation, and for integrating conventional as well as state-of-the-art technologies. Thus, it indicates that farther study is required on course re-evaluation, state-of-the-art technologies (e.g., mulsemedia) adoption challenges, the relationship between agility and e-learning adoption, and the status variations among schools (e.g., urban versus rural).

**Conclusion**

In general, our finding indicated that the experiences of teachers and students on e-learning resources is fair. However, there is only limited access to multimedia-rich resources, and premature practice of using them for the actual learning. The mapping of these results to the models for e-learning adoption also showed that the schools’ status correspond to the initial phase. Moreover, since furnishing schools with computing infrastructure and e-learning resources does not necessarily imply successful adoption, schools need to consider continued maintenance and upgrade; re-evaluate their curricula for the adoption; and aim for the next level in the selected e-learning adoption model. Accordingly, we proposed an agile adoption model. The implication of this study is that schools will be able to reduce the failure rates of their e-learning adoption initiatives due technological changes and unintended adoption. Our results also prompt further study on approaches of course re-evaluation for multimedia adoption; the impact of multimedia on students’ learning outcomes in terms of comprehension, creativeness, and problem solving; and the challenges of adopting edutainment, mulsemedia, virtual reality, gamification and other advanced technologies and concepts in terms of form factor, functionality and compatibility of the systems. In addition, our future research will also address more about the relationship between agility and e-learning, the status variations among schools (e.g., urban versus rural), and the progression in e-learning adoption in the upcoming years.
References


Panke, S., & Seufert, T. (2013). What’s educational about open educational resources? Different theoretical lenses for conceptualizing learning with OER. E-Learning and Digital Media, 10(2), 116-134.


Educational Games to Enhance Museum Visits for Schools

Benoît Bossavit¹, Alfredo Pina¹, Isabel Sanchez-Gil¹ and Aitziber Urtasun²

¹Public University of Navarre, Campus Arrosadia, Pamplona, Navarra, Spain // ²Fundación Museo Oteiza, Alzuza, Navarra, Spain // benoit.bossavit@unavarra.es // pina@unavarra.es // isabel.sanchezgil@unavarra.es // didactica@museooteiza.org

*Corresponding author

ABSTRACT

Museums usually look for new educational tools to enhance their exhibition. The Oteiza’s museum in Navarre (Spain) especially gives importance to the dissemination of Jorge Oteiza’s work to children at schools. Consequently, a didactics section was created with the objective of developing activities and relationship with schools. Jorge Oteiza represents one of the most important artists in the Spanish modern art and his sculptures stem from his proper philosophical concepts such as negative aesthetics via addition and subtraction, or activation of space and time. Such notions make the learning process at school complex. Thus, this study proposes a framework that aims to enhance the visit to the museum through a series of mini-games that shed light on these abstract concepts. Representative sculptures were selected and the corresponding activities were designed and developed in collaboration with the didactics section of the museum following a Co-Design approach. Then, the framework was tested by pupils from primary and secondary schools and students from educational practice. Therefore this paper provides a guideline to design educational games in collaboration with a museum, shows that mini educational games help students in learning artistic concepts and that motion-based touchless interfaces are not really adapted for classroom use.

Keywords

Virtual Reality, Museum, Interactive Learning Environments, Human-Computer Interface, Media in Education

Introduction

Museums acquire, conserve, research, communicate and exhibit cultural heritage for the purposes of education, study and enjoyment (ICOM, 2007). These institutions mostly target specialists, tourists and students (Bowen & Filippini-Fantoni, 2004) and thus, are usually looking forward to engaging their public in new types of displays in order to understand and appreciate the cultural heritage (Petridis et al., 2013). Museums do not limit their interests to indoor exhibits but also to the dissemination of the cultural heritage (Paliokas & Kekkeris, 2008). For instance, it has been shown that a clear learning strategy between schools and museums reinforce students’ learning experience (Griffin, 2004). The relationship between the entities can be strong with the creation of teamwork (students, educators, museum staff and researchers) which objective is usually to conceive specific museum activities (Wishart & Triggs, 2010). Such projects have shown interesting results, but it implies an additional cost for the schools and the stakeholders’ time schedule might invade their personal time (Vavoula et al., 2009). Another type of collaboration, which is not as strong as the abovementioned one, is the development of virtual museum applications specifically designed to be used at school (Paliokas & Kekkeris, 2008).

With such an objective, the Oteiza’s museum in Navarre (Spain), made a call for a project to disseminate Jorge Oteiza’s work at schools. This project, “Oteiza para tod@s” (literally, Oteiza for all), answered this call by proposing a framework that integrates the three following components: Art, Education and Technology. Jorge Oteiza represents one of the most important artists in the Spanish modern art (Alvarez-Martinez, 2003; Pelay, 1978). His sculptures move from at the vanguard art such as cubism or constructivism to his proper mathematical / philosophical concepts such as negative aesthetics (Alvarez-Martinez, 2003; Echeverria-Plazaola, 2012). Such concepts make the learning process at school complex and this is what motivates the didactics section of Oteiza’s museum to look for new educational tools (Urtasun, 2006).

Literature review

The “2015 Innovating Pedagogy report” pointed out that “crossover learning” would be one new learning strategy that may have deep influence in education (Sharples et al., 2015), aiming to combine formal and informal learning settings. For instance, one common approach that informal learning stems from is work-related tasks (Siemens, 2005) and active learning (Bonwell & Eison, 1991). This last consists in educational activities that engage physically students in the learning process, such as learning by doing things and thinking about the things they are doing (Bonwell & Eison, 1991). The gamification, for instance, is one mean to design such active educational activities. Therefore, this study aims to apply informal learning strategies in a more formal learning
setting with in and out of the school activities. In this sense, the theoretical framework of this project follows the “3T sandwich” model (Parsons, 2015) with three layers to help the elaboration of the educational tool: (1) the Theories that are used to develop and inform the educational tasks; (2) the Technology that supports learning and interaction; and (3) the Territories of use with the targeted users and the stakeholders including their views and experiences.

Regarding the Theories, several strategies have been set to develop Virtual Museum (VM) applications that aim at either: (1) delivering information to the visitors (Kuflik et al., 2011), (2) enriching the experience by focusing on the interaction between users and exhibits (Chen & Huang, 2012; Hsieh et al., 2014; Pescarin et al., 2013; Schieck & Moutinho, 2012; Şen, Díaz & Horttana, 2012) or (3) teaching specific content around a pedagogical task (Coenen, Mostmans & Naessens, 2013; Petridis et al., 2013). The objective of “Oteiza para tod@s” is to disseminate and to explain Jorge Oteiza’s concepts. Consequently, the future output belongs to this last category. The pedagogical tasks, in an informal learning setting, are usually integrated in game-based learning environments. Some evidence suggests learning efficiency (Clark, Tanner-Smith & Killingworth, 2016) due to the implicit use of cognitive skills such as memory or construction of knowledge (Sylaiou et al., 2009). These educational games, also known as serious games, adapt specific features of entertaining video games such as storyline or clear and short goals (Kapp, 2012), which make the experience enjoyable (Salen & Zimmerman, 2004).

There are two types of educational games, which have different aims: the mini-games and the complex-type ones. The complexity and richness of their corresponding features will influence the expected experience. For instance, complex games are usually long and combine various game mechanisms such as storyline, levels of difficulty, rewards or trophies to unblock. These latter are interesting for a deep and long-term learning process (Prensky, 2005). On the other hand, mini-games are usually short and do not require elaborated or evolutionary rules (De Jans et al., 2012; Prensky, 2005). The choice between these two types of games would depend mostly on the targeted audience, the place and the time allowed (Rivera-Gutierrez et al., 2014; Pescarin et al., 2013). For instance, visitors’ museums would sometimes prefer being engaged in the “experience” of learning rather than learning content by itself (Packer, 2006). Therefore, we decided to utilise pedagogic features of gamification that are known to support children’s learning such as short goals or scoring (Kapp, 2012). Gamification refers to “the use of game design elements in non-game contexts” (Deterding et al., 2011, p.9). A recent review about the increase of serious games in the field shows that about 29% of these aimed to an artistic market (Paliokas & Sylaiou, 2016). Amongst these studies, some educational games focused on engaging the visitors to interact with the museum artefacts either through quiz games (Belloti et al., 2012; Lien, 2015) or treasure hunting games (Rehm & Jensen, 2015; Yiamoutsou et al., 2014). The other approach found in the literature is to use storylines via adventure games to teach either art history with the different genres, eras and techniques (Froschauer et al., 2014; Pescarin et al., 2013). For future output belongs to this last category.

About the Technology, the current paradigms in Human-Computer Interaction (HCI) look for new means that propose engaging and intuitive interfaces via motion-based touchless interactions. These are also known as Natural User Interfaces (NUIs). One of the most common device that implements NUI is the Kinect, which is a device developed by Microsoft that detects 3D positioning and orientation of users’ skeleton. The Kinect is affordable, compact and easy to use. Since the release of the Kinect in 2011, research in HCI has actively looked for the use of body gestures to interact with virtual museum content. For instance, Rivera-Gutierrez et al. (2014) exposed in a museum of science a physical exhibit where visitors were able to learn about public health via mini educational games. However, the interaction with the system was not motion based but via pushing buttons. The Microsoft Kinect was used to detect the players. Yoshida et al. (2015) proposed a pedagogical mini game designed for children (10-12 years old) to support learning about paleontological era. The project is in its early stage and although the preliminary study showed positive insights of enhancing the interests of the learners, no evidence was shown that it was due to the interaction techniques. Mora-Guiard and Pares (2014) taught the concept of nanoscale to children (11-13 years old) via an exhibit that “miniaturises” the body over a huge surface of 10x4.5m. They showed that full-body experiences provided a better sense of scales to children. This, particularly shows that such a technology can be adopted by children from secondary school. Nevertheless, the system was set in a museum with a large open space, which is a very different environment from a classroom. Finally, targeting at a younger public (4-7 years old), Paul, Goh and Yap (2015) focused on promoting creativity and collaboration. The exhibition consisted on matching alphabet letters displayed on the screen with the projection of the body. The users’ participation and involvement revealed an interesting potential, although no formal evaluation was led. Again, this shows good insights that such a technology can also be adopted by children from primary school.
Concerning the Territories, the requisite of this project was to target students at school. The approach of this project is multidisciplinary (Art, Education and Technology) and thus, it was important to involve stakeholders in the design process of the future output. In this informal learning context, previous research has shown that the selection of the stakeholders and the type of collaboration depend mostly on the specific objectives of the project. For instance, when museums wanted to create new physical exhibits (Axelsen, Mygind & Bentsen, 2015; Dindler et al., 2010) or digital ones (Culén et al., 2013; Roussou & Ave, 2007), the research projects tended to engage typically developing children in Participatory Design (PD). These PD sessions typically consisted of an introduction to the corresponding museum field, followed by physical activities to elaborate ideas such as handicrafts (Culén et al., 2013; Taxén, 2004) or Lego plastic building blocks (Axelsen et al., 2015). On the other hand, when museums wanted to transfer knowledge (Dubois et al., 2011; Şen et al., 2012) or enrich the interaction between users and exhibits (Ciolfi et al., 2016; Coenen et al., 2013; Wishart & Triggs, 2010), the involvement of curators and ergonomists in User-Centred or Co-Design showed good results. Therefore, since the objective of this project was to convey and explain Jorge Oteiza’s artistic concepts at schools, we drew upon these good practices by engaging the head of the didactics section of the museum in a Co-design approach with several brainstorming sessions.

Research hypotheses

This paper presents a multidisciplinary project that aims to transfer artistic knowledge to children at school, through edutainment. To reach such an objective, we were interested in (1) the design process by taking into account the views and experiences of stakeholders from different background, (2) the usability at schools of advanced technological devices that promote motion-based interfaces and (3) the use of gamification in the learning process.

Overall, this project contributes to the field by discussing the impact of advanced technology and edutainment on learning and engagement. Additionally, the study was set in an authentic learning context. Indeed, the digital content was designed and developed over real sculptures and concepts of the artist and it was integrated during field trips to the museum, which was organised by the schools.

Thus, this project was built over the following hypotheses:
H1: Educational mini-games help students understand and learn artistic concepts.
H2: NUIs, and particularly motion-based interfaces, help students understand and learn artistic concepts.
H3: Educational mini-games are engaging.
H4: NUIs, and particularly motion-based interfaces, are engaging.

The following section describes the design process of the educational outcome. The framework is composed of four activities related to three sculptures / concepts of the artist. The digital tool was evaluated at three levels of education: primary school, secondary school and educational practice at University. Finally, the results are detailed and discussed.

Virtual museum application: Oteiza para tod@s

Design process

The Oteiza’s museum especially gives importance to the dissemination of J. Oteiza’s work to children. Consequently, the museum created a didactics section which is in charge of developing activities and relationship with schools. This project “Oteiza para tod@s” aimed to disseminate J. Oteiza’s work by combining Art, Education and Technology. Thus, to carry out such a multidisciplinary project several sessions were designed with stakeholders from different background: two researchers in Computer Science, one researcher in Art and Education, and the head of the didactics section of the Oteiza’s museum.

Session 1: Foundation

First, the researchers visited the museum and were guided by the head of the didactics section. The tour was directed according to the different periods of the artist’s life. Thus, the most representative sculptures of each artistic movement were thoroughly explained with their corresponding artistic style as well as the messages that
J. Oteiza wanted to transmit. The visit lasted approximately two hours. At the end of the session, each stakeholder was asked to list a number of sculptures that had aroused their interests.

Session 2: Selection of the content

Therefore, the objective of the second session was to select which sculptures and concepts would be digitalised. The session took place in the museum where the stakeholders shared their corresponding list with the reasons why they chose these sculptures. The head of the didactics section confirmed and corrected the researchers’ interpretation of the artist’s concepts. The selection of the final sculptures was decided by the common stakeholders’ interests, the technical feasibility estimated by the researchers’ point of view, and the potential added-value that VR can provide regarding the classic activities organised by the didactics section. Thus, the final selection was reduced to three sculptures.

Session 3: Idea generation

Finally, the last design session aimed to define the activities related to the corresponding concept of the sculptures. The stakeholders had previously decided to design mini-games because, contrarily to complex-games, a mini-game usually focuses on a single concept by using basic and easy-to-use game mechanisms (De Jans et al., 2017). Furthermore, mini-games facilitate the feeling of “challenge” with increasing levels of difficulties and the possibility of outperforming personal or other players’ score (Illanas et al., 2008). In that sense, Amr (2012) argues that challenges positively influence learning because the players feel a sense of achievement, which help them keep engaged. Serious games can be classified according to the Gameplay/Purpose/Scope model (Djaouti, Alvarez & Jessel, 2011). The scope refers to the market-based classification, which is in this study “Culture and Art.” The purpose is three-fold: broadcasting a message, training or exchanging data. The aim of this project is to transmit an educative message through the mini-games. Finally, the gameplay refers to the main features of the games. In other words, how the users interact with the game to fulfil the corresponding purpose. These games can be grouped in two categories: Game-based (Destroy, Match, Avoid, Block) and Play-based (Have luck, Write, Select, Move, Manage, Shoot, Create) (Djaouti et al., 2011).

Therefore, in order to optimise the session, the lead researcher (first author) proposed a scenario for each of the activities as a starting point. Each scenario defined a specific task to complete and elements of interaction. Then, for each activity, the ideas were discussed and adapted until all the stakeholders agreed on the main plot and features. At the end, all the stakeholders agreed on designing the mini-games with the game-based “Match” mechanism, and the play-based “Select, Move and Create” interactions. These mechanisms fit the main concepts of the artist, which are mostly based on the way the sculptures were created and their relationship with the environment regarding the point of view.

The mini games

The software was developed using C#, the Microsoft Kinect SDK and GoblinXNA as rendering engine. It was decided that the games would not diffuse any sound because it can be disruptive in both school and museum environment (Economou, 1999).

Negative aesthetic via subtraction

The first activity was inspired of a permanent exhibit which is exposed in Pamplona, Spain. The sculpture had been built by removing spherical matters which resulted in a two-column shape (see Figure 1). The concept behind this sculpture, however, is the negative aesthetic. Therefore, it is important to focus the attention not only on the sculpture but also on the invisible matter generated by this latter. In order to explain this first concept, the mini game moves automatically the users’ point of view around a virtual reproduction of the sculpture (not strictly identical to the original one). Then, the objective is to find out of four possibilities the empty space that matches the invisible matter (see Figure 1). The player must analyse the scene from the specific point of view, and decide the option with the “Select” feature. To select an option, the users should point at the screen and move the projection of the hand towards the corresponding icon at the top of the screen. If the player fails, a cross is drawn on the wrong selection as feedback and the player can still select another option, however, the
final score will decrease. Once the solution has been found, the invisible matter is highlighted and turns into spherical shapes to explain the real reasoning of the artist. The gameplay feature is limited to selection to let the player all necessary time to observe and understand the scene with a reduced cognitive load.

Figure 1. On the left, picture of the real sculpture. On the right, screenshot of the corresponding activity

Negative aesthetic via addition

Figure 2. On the left, picture of the real sculpture. On the right, screenshot of the corresponding activity

Activation of space and time

The second activity was based on a sculpture created by addition of matter (see Figure 2). The concept behind this work is also negative aesthetic, although this time the invisible matter is not spherical but cubic. Since this sculpture was built by addition, the mini game offers similar mechanisms. The position and orientation of half of the piece is set in the scene with its invisible matter highlighted in green (light grey in Figure 2). The user manipulates the second part of the sculpture by using the Crank Handle technique (Bossavit et al., 2014). This is a one-hand manipulation technique were the movement of the hand is transferred to the virtual object and its orientation is set with a circular movement of the hand around one of the three primary axis as if the user were turning a crank handle. The corresponding invisible matter of the second piece is highlighted in blue (dark grey in Figure 2) to its final position within the final sculpture (see Figure 2). Thus, the user has to position and orientate this second half of the sculpture in order to encapsulate its corresponding invisible matter (highlighted in blue) while being in direct contact with the static bit. This final position will shape one and unique sculpture. The feature of the game is, then, limited to “Move,” in order to let the user reproduce the construction of a sculpture by taking into account the environment. In that sense, collisions are calculated so that the two pieces cannot overlap. Furthermore, an extra magnet feature is activated when the two pieces are closed to the solution, and thus, the manipulated piece is automatically placed to match the fixed one. Several sculptures are designed so that users will not perform the same one twice in a row. The challenge is increased by adding time to the task completion. Indeed, the task should be completed within a time range limit (40 seconds) otherwise the solution would be animated once the time is over.
The third activity was based on a sculpture for which J. Oteiza had found inspiration in Pieter Mondrian’s work. J. Oteiza represented the activation of the space by curving Mondrian’s parallel and perpendicular lines which resulted in an opened spherical shape (see Figure 3). It is important to state that the sphere is only defined by its negative aesthetic. Furthermore, the concept behind this work remains somehow open to interpretation because, and in accordance with the didactics section of the Oteiza’s museum, the strips would not only represent the surface of the negative sphere. Indeed, each strip would also stand for a “negative” sphere at one specific time. As a result, the mini game was designed with two stages. During the first phase, the user is presented several “flat” patterns and has to find which one fits the model (see Figure 3). At this stage, the player is offered the “Select” feature to choose the correct model. The user can manipulate a virtual ray, which is the extension of the arm, by pointing at the screen. Once the virtual ray intersects a starting point of a strip, which is represented by a small sphere, the user can activate the concept of space by closing the hand. Thus, an animation, which shows how the “flat” pattern converts into the final sculpture, would be launched. In case of wrong answer, the generated sphere comes back to its original flat shape, and the player loses a life and is given a new opportunity. Several levels are designed to increase the difficulty. The first level has one “flat” pattern with several starting points. The second level has three patterns with only one starting point each. Afterwards, the level has three patterns with several starting points. Once the player loses all the lives or reaches the last level, the strips of the generated sphere are animated around the negative sphere. Each strip moves independently at different speeds. Then, the user is offered the “Move” feature to reconstruct the original sculpture by rotating the strips one by one using the Crank Handle technique (Bossavit et al., 2014). Therefore, this activity proposes both the “Select” feature to have time to observe the scene and the “Move” feature to feel the concept behind the construction of the sculpture. The challenge increases with levels and is enhanced with a limited numbers of lives.

Pilot study

Once the three activities were implemented, a session was set in the Oteiza’s museum with the vice director of Oteiza’s museum and three professors in art history (two from the University of Barcelona in Spain and one from the Public University of Navarre in Spain). The session was organized in three stages. First, all the stakeholders visited the museum (approximatively one hour). Then, the researchers explained the different activities and how to interact. The stakeholders were given the possibility to interact as well (about one hour). At the end, the participants provided feedback in order to improve the representation of the activities (reported in the “Findings and Discussion” section).

Furthermore, they found that activities all lacked the presence of creativity. This was also suggested by Kuo et al. (2009) who called for adding artistic teaching materials into E-learning programs. Thus, they suggested that users should be able to create their own sculpture based on the concept outlined by the software. Consequently, we added the “Create” feature within a new activity that allows users to add and edit pieces by changing their size (length, width and thickness), their curvature and their position / orientation. Users can manipulate the pieces separately or together. They are also offered the possibility to observe the invisible matter that the sculpture creates (see Figure 4). The 3D manipulation is done via the Crank Handle technique (Bossavit et al., 2014) and the parameters that change the shape of the piece are controlled by sliders. We proposed two techniques to manipulate the sliders: (1) via a constant step by touching the shoulders with the hand; (2) via an interpolated value. For this second technique, an unfolded user’s arm stands for the slider while the other hand represents the tick between the hand and the shoulder (Shoemaker et al., 2010).
Figure 4. Screenshot of the last activity which focuses on the creation of a sculpture that is based on the concepts learnt from the other three activities

Evaluation at schools

This evaluation study seeks for the impact of technological input on the illustration of artistic concepts at schools. However, the design and development of this educational tool does not intend to replace the visit of the museum but to enhance it. Consequently, comparing the impact of this tool as an alternative to the visit at the museum is beyond the scope of this evaluation.

Participants

Schools usually organise field trips to museums from primary to University. Therefore, participants from different ages and educational stages had been recruited for this study. The first group was composed of 57 primary pupils from a state school of Pamplona, Spain. Participants’ mean age was 7.14 ($SD = 0.43$) which included 21 male and 36 female. 28 participants played with the software (active) and 29 just watched (passive). The second group was composed of 60 secondary pupils from another state school of Pamplona, Spain. Participants’ age was 14.55 ($SD = 0.58$) which included 24 male and 36 female. 19 participants played with the software (active) and 41 just watched (passive). The last group was composed of 21 students in educational practice at the Public University of Navarre, Spain. Participants’ mean age was 22.1 ($SD = 2.67$) which included 2 male and 19 female. 13 participants played with the software (active) and 8 just watched (passive).

Procedure

Each group followed the same procedure. None of the educational institutions teaches content about J. Oteiza in their programme, thus, according to the teachers, the class before the visit to the museum was an introductory session about the artist.

Afterwards, the pupils and students visited the Oteiza’s museum during approximately one hour. The visit was guided by the head of the didactics section of the museum. The tour started with a presentation of J. Oteiza who is a reference as artist and intellectual of the twentieth century. Then, the visit continued on the permanent exhibits with the most representative sculptures. Not only does Oteiza’s work open to specific concepts such as figurative/abstract art, movement/space, or visible/non-visible matter, but it also refers to social art and sculpture-architecture relationship.

Later, within a range of one to four weeks after the visit of the museum (according to the availability of the schools), the researchers were contacted to evaluate the use of the software with the students during one of the mainstream teaching session at the school. Thus, we setup two installations in the same room so that participants had more opportunities to play. The game was projected onto a wall with the Kinect under it and the interacting student was placed at about 2 meters from the screen (see Figure 5). The session lasted 50 minutes during which the students could voluntarily test one out of the four activities for about 40 minutes in total. The researchers were there to help the users interact if required. For the last 10 minutes, the participants were asked to fill questionnaires (see further details below).
Results

This study is based on the hypotheses that the targeted users’ experience, through mini games and motion-based interfaces, help understand and learn artistic concepts (H1 and H2), and are engaging (H3 and H4). Therefore, we used two different questionnaires to measure these statements, and for all the results we looked different independent variables, i.e., Gender (whether participants were (1) Male or (2) Female); Device (whether participants interacted with the software: (1) Active or (2) Passive interaction); and Education (whether participants are from (1) Primary, (2) Secondary or (3) Educational Practice).

Two-tailed independent samples t-tests were performed on the participants’ scores for the two independent variables Gender and Device. And one-way ANOVAs were performed for the comparison between the three groups of participants regarding Education. Pairwise comparisons adjusted by a Bonferroni correction were applied when significant differences appeared.

Learning experience

The participants’ learning experience was surveyed with a multiple choice questionnaire composed of ten questions with four possible answers each (see Table 1). The first fifth questions (part I) were related to the author himself and the second half (part II) quizzed about the Oteiza’s concepts treated by the software. The questions were elaborated by the head of the didactics section of the museum and it was given out at the pre-visit session and after testing the mini-games. The “post-test” was the same as the “pre-test” although the answers were inverted.

Table 1. Questionnaire that was used to evaluate the learning experience (answers are highlighted in italic)

<table>
<thead>
<tr>
<th>Category</th>
<th>Questions</th>
<th>Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning Oteiza</td>
<td>1. Where is Jorge Oteiza born?</td>
<td>San Sebastián / Pamplona / Orio / Bilbao</td>
</tr>
<tr>
<td></td>
<td>2. In the field of Art, Oteiza was well-known for?</td>
<td>Moviemaker / Sculptor / Painter / Architect</td>
</tr>
<tr>
<td></td>
<td>3. In 1935 the artist had left Spain for another country in which he had</td>
<td>France / Germany / South America / Italy</td>
</tr>
<tr>
<td></td>
<td>been working as teacher and artist until 1948. Where was it?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. Which is the city that owns the most work of Oteiza within the streets?</td>
<td>Pamplona / San Sebastián / Biarritz / Madrid</td>
</tr>
<tr>
<td></td>
<td>5. The Oteiza’s museum was designed by an architect from Navarre. Who was</td>
<td>J. Sáenz de Oiz / F. Mangado / R. Moneo / V. Eusa</td>
</tr>
<tr>
<td></td>
<td>he?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7. Which context is the most important when one is studying Oteiza?</td>
<td>Movement / Shape / Empty space / Color</td>
</tr>
<tr>
<td></td>
<td>8. Which one is the most related to space</td>
<td>The sculpture fills a space /</td>
</tr>
</tbody>
</table>
No direct relationship between sculpture and space
The empty space means nothing to the sculpture
The empty space can become a sculpture

9. How many dimensions can be observed in a sculpture?
Two / one or two / three / two or three

10. Does the color mean something to the sculpture?
Yes, since it changes its aspect
No, it is only important in painting
No, it is only important in painting and drawing
Only if the sculpture is figurative

In order to analyse the learning gains, dependent variables were defined:
- \( \text{Learning} = \left( \frac{\text{post-test} - \text{pre-test}}{10} \right) \times 100 \)
- \( \text{Learning}_\text{Oteiza} = \left( \frac{\text{post-test}_\text{partI} - \text{pre-test}_\text{partI}}{5} \right) \times 100 \)
- \( \text{Learning}_\text{Concepts} = \left( \frac{\text{post-test}_\text{partII} - \text{pre-test}_\text{partII}}{5} \right) \times 100 \)

The results of the statistical analysis over this questionnaire are summarised in the Table 2. The dependent variable “Questionnaire” represents the score over 10 points. Each correct answer was marked one point. The three abovementioned dependent variables “Learning,” “Learning_Oteiza” and “Learning_Concepts” represent the percentage of increased knowledge between the “pre-test” and the “post-test.” The independent variables were not reported when there were no significant differences.

<table>
<thead>
<tr>
<th>Group</th>
<th>Dependent variable</th>
<th>Independent variable</th>
<th>Mean</th>
<th>SD</th>
<th>Results</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>Questionnaire</td>
<td>Pre-test</td>
<td>2.57</td>
<td>1.55</td>
<td>t(56) = 8.8</td>
<td>***p &lt; .001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-test</td>
<td>4.49</td>
<td>1.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Learning</td>
<td>Male</td>
<td>16.6</td>
<td>17.12</td>
<td>t(55) = -0.862</td>
<td>p &gt; .05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female</td>
<td>20.5</td>
<td>16.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Active</td>
<td>14.6</td>
<td>13.7</td>
<td>t(55) = 2.08</td>
<td>*p &lt; .05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Passive</td>
<td>23.4</td>
<td>17.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Learning_Oteiza</td>
<td>Active</td>
<td>17.14</td>
<td>22.25</td>
<td>t(55) = 0.76</td>
<td>p &gt; .05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Passive</td>
<td>22.06</td>
<td>26.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Learning_Concepts</td>
<td>Active</td>
<td>12.14</td>
<td>19.88</td>
<td>t(55) = 2.29</td>
<td>*p &lt; .05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Passive</td>
<td>24.82</td>
<td>21.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary</td>
<td>Questionnaire</td>
<td>Pre-test</td>
<td>4.88</td>
<td>1.51</td>
<td>t(59) = -9.71</td>
<td>***p &lt; .001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-test</td>
<td>6.98</td>
<td>1.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Learning</td>
<td>Male</td>
<td>14.58</td>
<td>16.14</td>
<td>t(58) = 2.53</td>
<td>*p &lt; .05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female</td>
<td>25.27</td>
<td>15.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Active</td>
<td>18.42</td>
<td>18.63</td>
<td>t(58) = 0.81</td>
<td>p &gt; .05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Passive</td>
<td>22.19</td>
<td>15.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Learning_Oteiza</td>
<td>Male</td>
<td>19.16</td>
<td>16.12</td>
<td>t(58) = 1.35</td>
<td>p &gt; .05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female</td>
<td>25.55</td>
<td>18.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Learning_Concepts</td>
<td>Male</td>
<td>10</td>
<td>26.34</td>
<td>t(58) = 2.22</td>
<td>*p &lt; .05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female</td>
<td>25</td>
<td>24.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Educational Practice</td>
<td>Questionnaire</td>
<td>Pre-test</td>
<td>7.00</td>
<td>1.51</td>
<td>t(20) = 3.11</td>
<td>*p &lt; .01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Post-test</td>
<td>8.19</td>
<td>1.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Learning</td>
<td>Male</td>
<td>10</td>
<td>14.14</td>
<td>t(19) = -0.15</td>
<td>p &gt; .05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female</td>
<td>12.1</td>
<td>18.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Active</td>
<td>11.53</td>
<td>18.63</td>
<td>t(19) = -0.11</td>
<td>p &gt; .05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Passive</td>
<td>12.5</td>
<td>16.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inter groups</td>
<td>Learning_Oteiza</td>
<td>Primary</td>
<td>19.64</td>
<td>24.34</td>
<td>F(2,135) = 0.4</td>
<td>p &gt; .05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Secondary</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Educational practice</td>
<td>23</td>
<td>17.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Learning_Concepts</td>
<td>Primary</td>
<td>18.59</td>
<td>21.66</td>
<td>F(2,135) = 3.38</td>
<td>*p &lt; .05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Secondary</td>
<td>19</td>
<td>26.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Educational practice</td>
<td>3.8</td>
<td>24.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Learning</td>
<td>Active</td>
<td>15.16</td>
<td>16.41</td>
<td>t(136) = 2.27</td>
<td>*p &lt; .05</td>
</tr>
</tbody>
</table>
Passive 21.66 (%) 16.78
Male 15.31 (%) 16.26 \( t(136) = -1.77 \) \( p > .05 \)
Female 20.65 (%) 16.98

Note. \(^*p < .05; \(^{**}p < .001.\)

Engagement

The participants’ experience was surveyed via a 5-Likert scale questionnaire with 12 questions. Four questions were related to motivation (1-4), four about usability (5-8) and four about utility (9-12). The questionnaire was given out at the end of the session after testing the mini-games (see Table 3).

Table 3. Questionnaire that was used to evaluate engagement and motivation

<table>
<thead>
<tr>
<th>Category</th>
<th>Question</th>
</tr>
</thead>
</table>
| Motivation | 1. I like going to museums  
2. When I am in a museum I usually watch the exhibits  
3. When I am in a museum I like spending time with the interactive exhibits  
4. I enjoyed the experience and learnt about Oteiza’s work |
| Usability | 5. I found it easy using the application  
6. I found it difficult understanding what I had to do  
7. I could use this application without the help of an adult  
8. I found it difficult completing the activities |
| Utility | 9. I think the application helped me understand better J. Oteiza’s work  
10. I think the visit to the museum is enough to understand J. Oteiza’s work  
11. I would like such an application in my school  
12. I would go more often to museums if such kind of applications would be set up |

The analysis of internal consistency reliability of the questionnaire revealed a Cronbach’s alpha coefficient at 0.695. The Cronbach’s alpha coefficient ranges between 0 and 1 and an acceptable minimal reliability value is 0.7 (Nunnaly, 1978). The results of the statistical analysis on the three dependent variables: Motivation, Usability and Utility are summarised in the Table 4. The score displayed in the “Mean” column represents the sum of the answers of the four questions corresponding to the dependent variable. The mark is over 20.

Table 4. Statistic results of the questionnaire about engagement (significant differences are marked with an asterisk)

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Independent variable</th>
<th>Mean</th>
<th>SD</th>
<th>Result</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivation</td>
<td>Male</td>
<td>13.91</td>
<td>3.18</td>
<td>( t(134) = -3.66 )</td>
<td>(^*p &lt; .05)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>15.77</td>
<td>2.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Active</td>
<td>15.81</td>
<td>2.68</td>
<td>( t(134) = 2.4 )</td>
<td>(^*p &lt; .05)</td>
</tr>
<tr>
<td></td>
<td>Passive</td>
<td>14.61</td>
<td>3.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Primary</td>
<td>*16.5</td>
<td>2.98</td>
<td>( F(2,133) = 15.42 )</td>
<td>(^{**}p &lt; .001)</td>
</tr>
<tr>
<td></td>
<td>Secondary</td>
<td>*13.76</td>
<td>2.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Educational practice</td>
<td>15.31</td>
<td>2.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usability</td>
<td>Male</td>
<td>12.04</td>
<td>3</td>
<td>( t(134) = -0.63 )</td>
<td>( p &gt; .05 )</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>12.38</td>
<td>2.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Active</td>
<td>12.57</td>
<td>3.28</td>
<td>( t(134) = 1.06 )</td>
<td>( p &gt; .05 )</td>
</tr>
<tr>
<td></td>
<td>Passive</td>
<td>12.02</td>
<td>2.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Primary</td>
<td>12.12</td>
<td>3.31</td>
<td>( F(2,133) = 0.87 )</td>
<td>( p &gt; .05 )</td>
</tr>
<tr>
<td></td>
<td>Secondary</td>
<td>12.6</td>
<td>2.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Educational practice</td>
<td>11.63</td>
<td>2.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utility</td>
<td>Male</td>
<td>13.55</td>
<td>2.95</td>
<td>( t(134) = -2.28 )</td>
<td>(^*p &lt; .05)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>14.7</td>
<td>2.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Active</td>
<td>14.88</td>
<td>2.1</td>
<td>( t(134) = 2.07 )</td>
<td>(^*p &lt; .05)</td>
</tr>
<tr>
<td></td>
<td>Passive</td>
<td>13.87</td>
<td>3.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Primary</td>
<td>*14.98</td>
<td>2.39</td>
<td>( F(2,133) = 3.31 )</td>
<td>(^*p &lt; .05)</td>
</tr>
<tr>
<td></td>
<td>Secondary</td>
<td>*13.65</td>
<td>3.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Educational practice</td>
<td>14.36</td>
<td>2.56</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Findings and discussion

This project aims at disseminating the artistic concepts of J. Oteiza to children at school. Thus, the adopted learning strategies mixed formal and informal learning by using interactive gaming activities to promote individual learning. In this sense, we hypothesised that educational mini-games and motion-based interfaces would help participants understand and learn abstract artistic concepts and being engaging.

Research hypotheses

In terms of the impact of the educational mini-games in the learning process, the scores of the questionnaires show that all the groups increased their knowledge significantly (see Table 2). The overall learning experience aimed to provide information about the artist (mostly from the museum visit) as well as his artistic concepts (mostly from the mini-games). The increase of knowledge about J. Oteiza is similar to all the groups (between 19.6% and 23.0%, see Table 2). This means that the visit of the museum was efficient to everyone. On the other hand, the increase of knowledge about the artist’s concepts varied depending on the educational group. The pupils from secondary school learnt the most (increase of 19%, see Table 2), followed by the ones from primary school (increase of 18.5%, see Table 2). This puts forward the fact that children from 7–8 years are, indeed, able to learn complex concepts as suggested by Antoniou et al. (2013). However, the students from educational practice showed a very small progression (increase of 3.8%, see Table 2). This is probably due to the fact that concepts such as invisible matter and spatiotemporal dimensions are already assimilated by adults.

Concerning the impact of the technology, we compared the scores obtained by the users who interacted with the software and the one who only saw the others interacting. People’s behaviour and expectation regarding digital exhibits might vary from the environment. For instance, in museums, visitors tend to understand and learn through active participation (Kampouroupolou et al., 2013). However, in this study, students who did not interact directly with the software learnt significantly more (increase of 21.6% against 15.1%, see Table 2). An explanation of such results might be humans’ social boundaries issues. Indeed, the students interacted in front of their colleagues, which might be a very stressful experience due to the importance we give to the judgment of people that surround us (Feinstein, 2004). Thus, most of the students’ attention might have focused on the way they behaved instead of the task by itself. Furthermore, the stress is higher when people are socially closer. In museums, this pressure is lowered and some people manage to negotiate these social boundaries (Schick & Moutinho, 2012; Paul et al., 2015). Another reason that could explain why the students learnt more when they were not active is the cognitive load that was required by the interface. Indeed, all the participants did not score high usability of the system (score of 12 out of 20, see Table 4), which means that they found it relatively difficult. The activities were designed with the Crank Handle technique (Bossavit et al., 2014), which allows manipulation of six degrees of freedom (three for translation and three for rotation). Although the technique is easily understandable because it is based on a common metaphor, which is rotating a crank handle, the accuracy of the Kinect requires users to perform clear gestures. Thus, this technique does entail some training that the children did not have. Consequently, it might be judicious to provide more assistance to the users by limiting the complexity of the gestures (Pescarin et al., 2013; Yoshida et al., 2015) as well as the amount of actions (Hsieh et al., 2014; Mora-Guiard & Pares, 2014).

In terms of engagement, besides improving significantly their understanding of the abstract notions, the pupils from primary school revealed the highest rate of motivation (score of 16.5 out of 20, see Table 4) and usefulness (score of 14.9 out of 20, see Table 4). Thereafter, the primary school teacher commented that the session had motivated her and she organised a workshop where children could sculpt and collage works related to J. Oteiza’s concepts (see Figure 6). Surprisingly, the teenagers, who increased the most their knowledge, revealed the least motivation (score of 13.7 out of 20, see Table 4) and just found the experience somehow useful (score of 13.6 out of 20, see Table 4). This contrast has already been observed in museums where the enjoyable experiences and the amount of cognitive learning varied regarding the group of participants (Griffin, 2004).

About the impact of the technology, the participants who interacted with the software revealed a higher motivation (score of 15.8 against 14.6 out of 20, see Table 4) and a higher score in utility (score of 14.8 against 13.8 out of 20, see Table 4). This goes in line with the literature that also revealed evidence of this type of interfaces being engaging (Hsu, 2011; Lee et al., 2012).
Theoretical framework

Nevertheless, all these results suggest some important recommendations for future practitioners and academics regarding the elaboration of the theoretical framework. This study was grounded in the “3T sandwich” model (Parsons, 2015) that focuses on three aspects: the Theories (in this project the gamification), the Technology (motion-based interfaces) and the Territories of use (students as target and multidisciplinary stakeholders as designer).

Regarding the gamification, beside the fact that mini games are engaging and help increase knowledge, another interesting outcome is the gender effect. Indeed, female participants were the most engaged, they found the experience more motivating than male participants (score of 15.7 against 13.9 out of 20, see Table 4) and more useful (score of 14.7 against 13.5 out of 20, see Table 4). This might be explained by the fact that girls tend to prefer games that require cognitive skills instead of destruction-like games (Pasek, 2008; Pelletier, 2008). Furthermore, it appears that female participants increase the most their knowledge (20.6% against 15.3%, see Table II) and especially in secondary school (25.2% against 14.5%, see Table 2). This suggests that gender issues should be taken into account when designing educational games in order to adapt the content to the main preferences of the participants (Holtzblatt & Kules, 2017; Kafai, 2008).

In terms of technology, this study shows that albeit motion-based interfaces are engaging, these kinds of devices are not well adapted in a school context. Indeed, it is limited to one or two players at the same time, which can affect social boundaries regarding the others students. Additionally, it also requires an open space without occlusion in order to avoid interferences.

Finally, concerning the design methodology, the results of the pilot study showed that the collaboration with the didactics section of the museum was indeed efficient since at the first try, all the experts agreed that the activities did explain J. Oteiza’s abstract concepts. However, we do recommend an extra iteration with different experts to ensure that the key message is conveyed correctly. Indeed, the stakeholders’ experience in pedagogy and art revealed important improvements for the interpretation of the abstract concepts. For instance, the stakeholders considered that the tutorials, which explained how to interact with the software, were represented the same way as the pedagogical information. Therefore, they suggested replacing the text by simple animations, which apart from being more visual would definitely separate interactions from pedagogical aspects (Antoniou et al., 2013). Another interesting aspect is that all the stakeholders agreed that it would be interesting to add new quotations of J. Oteiza within the games so that it would impregnate the user with Oteiza’s personality. Finally, they felt like the objective of the game was to replace the visit of the museum. As a result, they accentuated that this digital tool should be a complement to the visits, which was also suggested by Sylaiou et al. (2009). This is an important aspect that should be taken into account because it affects the future design of the educational tools.

Limitations and recommendations

Although, the design and development of this educational tool did not intend to replace the visit of the museum but to enhance it, the presence of a control group could improve consequently the data analysis by quantifying the different impact on the final learning progress from the visit of the museum and the use of the mini-games.
Another limitation of this study is the qualitative questionnaire (see Table 3). It was designed by the researchers and its reliability Cronbach’s alpha coefficient was at 0.695 when an acceptable rate is 0.7. This means that, strictly speaking, the questionnaire may be questionable and this let the readers decide about its relevance.

Anyhow, the current study revealed some interesting limitations that can be avoided in further studies. For instance, motion-based interaction games at school deal with humans’ social boundaries issues, which has a direct impact on the learning process. Therefore, it might be more judicious to either (1) design games that requires a more anonymised interactions by using more classical interfaces like mouse and keyboard, or (2) organise individual or open sessions where the students can interact when they feel more confident.

Furthermore, using motion-based interaction might require a certain cognitive load, which may hinder the learning process. Thus, we recommend providing more assistance to the users by limiting the complexity of the gestures.

Conclusions

A series of mini games that shed light on Jorge Oteiza’s artistic concepts to children have been co-designed with the didactics section of a museum. In informal learning settings such as Art Museums, the educational content are strictly related to specific concepts / artefacts, which make the outcomes poorly replicable. That is why, in this study, the emphasis was also put on the design methodology based on the 3T sandwich model. The framework was evaluated by children from primary and secondary schools and students from educational practice in order to validate the design of the educational mini-games using of Natural User Interfaces. Thus, both the analysis of the users’ experience towards the technology and their learning experience helped us provide some recommendation to future practitioners in order to ground the theoretical framework. For instance, about learning outcome, the participants did learn more about Jorge Oteiza as well as his artistic concepts, including children from primary school. This shows that the designed educational tool completes its objective efficiently. Regarding the targeted users, the study highlights the interesting fact that girls were, in fact, more engaged and learnt the most. This might be due to the game mechanisms which focused more on cognitive than destruction-like skills. Besides, it was observed that, contrary to visits at museums, participants who did not interact were more likely to understand and increase their knowledge. This means that an informal tool based on active learning may have another effect within a formal learning context. This might be a consequence of using Natural User Interfaces at school, where the, participants had to interact alone in front of all their friends, which obliged them to negotiate their social boundaries beyond shyness and stress. This side effect might be reduced by limiting the interaction to simple gestures so that participants can focus more on the educational content.

Overall, we argue that applying informal educational learning tools within a formal learning context enhances Museum experiences and link them to classrooms, which promotes crossover learning. However, this is a short-term study of an overall project that deals with advanced technologies, on the one hand, and learning strategies, on the other hand. Therefore, it would also be interesting to complement this study with two additional evaluations. For instance, one study should focus on the real impact of advanced technologies at school by using this same framework but with different technology. The other study should specifically focus on the learning strategies comparing formal educational tools with this game-based learning tool using a control group.

Acknowledgments

We would like to thank the KUTXA OBRA SOCIAL who funded this project and the Oteiza’s Museum for their availability and help in carrying out this work. Our acknowledgment goes to all the students who participated in the evaluation as well as their respective schools: IES Barañain with Elena Medrano Muñoz; Camino de Santiago en Zizur Mayor with Marta Aguas; Calasanz-Escolapios in Pamplona with Carlos Jimenez and the Public University of Navarre with Amaia Arriaga Azcarate. Benoît Bossavit had a grant from the Universidad Pública de Navarra, Spain.

References


Using Exaggerated Feedback in a Virtual Reality Environment to Enhance Behavior Intention of Water-Conservation

Wei-Che Hsu¹, Ching-Mei Tseng² and Shih-Chung Kang³
¹Department of Civil Engineering, National Taiwan University, Taipei City, Taiwan R.O.C. // ²Department of International and Comparative Education, National Chi Nan University, Nantou County, Taiwan R.O.C. // ³Department of Civil and Environmental Engineering, University of Alberta, Alberta, Canada

*Corresponding author

ABSTRACT
Taiwan has long been categorized as a country suffering from water shortages. The acquisition of knowledge and development of the concept of water conservation, and the development water conservation habits, are considered a civic responsibility of the citizens. The purpose of the present study was to utilize an immersive virtual environment technology (IVET) to create virtual experiences that expose individuals to vivid information with personal relevance and immediacy in hopes of increasing the behavior intention to conserve water. A mediated virtual experience was provided in the form of an experiential learning game, in which participants were situated in a virtual bathroom and asked to repeatedly use a 600-milliliter bottle to fill a specific water tank to flush a toilet and take a one-minute shower. While consuming the water resources, participants received exaggerated feedback (EF) to intensify the negative consequences of water consumption (direct EF) and/or environmental damage (ambient EF) that emphasized personal affective responses. A total of 165 players, separated into four groups according to the ambient or direct EF conditions, experienced the game activities. ANOVA was used to examine the effects of the experimental intervention. The results showed that the immersive virtual reality game in this study caused significant changes in cognition and behavior intention. This study contributes a novel persuasive technology specific to water resources.

Keywords
Water Environmental Education, Water conservation, Virtual reality, Experiential learning, Exaggerated feedback

Introduction
Long categorized as a country suffering from water shortages, Taiwan has only about 4,074m³ of annual rainfall per person, less than 20% of the world average of 21,796m³. The government and industry have sought to promote water resource conservation activities, and students and the general populace are expected to learn about water conservation. Recent water conservation education in Taiwan has focused on theoretical and conceptual knowledge. Although the education, in the form of knowledge domains and procedures, and effectiveness and social knowledge, is still able to influence students’ pro-environment attitudes (Larson & Redman, 2014; Steg & Vlek, 2009), a particular challenge is that the environment-related knowledge is abstract, is comprised of multiple behaviors, and has low personal relevance, thus causing limitations in the individual’s ability to distinctly perceive the environmental issue (Bray & Shackley, 2004; Pooley & O’Connor, 2000; Redman, 2013; Swaim et al., 2014). Limited consciousness of the issue continues to be an impediment to automatic triggering of individual environmental responsibility and behavioral modification (Breunig et al., 2014; Mobley et al., 2010).

Awareness of the environmental problem is an important prerequisite for promoting pro-environmental behaviors (De Groot & Steg, 2009; Grothmann & Patt, 2005; Van Der Werff & Steg, 2015). However, the three major characteristics of environmental risk make it more challenging to achieve personal higher perceptual and behavioral responses. First, the use of the natural resources is abstract. This abstraction leads to cognitive complexity, which causes people commonly to underestimate the actual amount of resource consumption (Attari et al., 2010; Jamieson, 2006). When the misconception is constructed in individuals’ minds, they lack motivation to review their pro-environmental notions and habits (Markowitz & Shariff, 2012). Second, environmental degradation, like the scarcity of water, is hard to observe directly and immediately (Weber, 2006). This uncertain future consequence of the environment creates little personal concern for water as a scarce resource and lowers people’s intention to care about the water conservation (Moser, 2010; Newton & Meyer, 2013; Uzzell, 2000; Wiek et al., 2011). Individuals then consider the pro-environmental behavior as an unnecessary adaptation. Finally, the daily consumption of water and energy is unnoticeable. With the insulation of modern people from the environment, individuals cannot easily perceive the excessive resource usage of their personal habitual behaviors (Blühdorn, 2011). Therefore, despite people’s belief in the existence of environmental problems, they...
still feel less responsibility to make efforts to develop conservation behaviors (Kagawa, 2007; Parker & Sams, 2015; Sarabia-Sánchez et al., 2014).

In summary, the temporal distance between an unnoticeable cause and the abstract effect of environmental problems causes difficulties in raising environmental consciousness. To reduce the gap and further engage people in pro-environmental behaviors, it is necessary to portray water consumption vividly and conceptualize the environmental risk as the negative consequences of personal behaviors. Indeed, it is critical that individuals be provided with direct experiences that effectively clarify and increase individual perceptions of the problem.

In the current study, the authors aimed to develop an effective intervention for individuals to increase awareness of personal relevancy (cause) and vivid consequences (effect) through a persuasive experience simulated by immersive virtual environment technology (IVET). We used IVET to produce a virtual reality game that illustrates how individuals directly cause negative consequences as they consume water. We predicted that a vivid and personally relevant virtual experience (e.g., depleting one’s physical energy to afford habitual water usage) would trigger individual cognition, change the attitude toward the water conservation issue, and increase behavior intention to conserve water. Furthermore, we proposed the incorporation of immediacy by speeding up the time to display accelerated negative outcomes in the experiential feedback of consuming water resources, known as exaggerated feedback (EF). Participants were exposed to a scenario wherein when they consumed water, the loss of water resources and the degradation of the external environment would be aggravated; i.e., the actual feedback was exaggerated. We assumed that the virtual experience applying exaggerated feedback would result in higher levels of cognitive and affective responses and produce greater improvement of behavior intention to conserve water. Specifically, two research questions were examined as follows:

Q1: Does experiencing vivid and personally relevant virtual experience change participants’ responses in water saving cognition, attitude, and behavior intention?

Q2: Does the exaggerated feedback in the virtual experience have a positive impact on water saving cognition, attitude, and behavior intention?

Vivid information of the effect

In resource conservation policies and programs, the vivid effect, or solid abstract information, can significantly accelerate individual understanding of the actual consumption and make messages more persuasive (Dillard & Main, 2013; Sheppard, 2012). Research suggests that the strategy of communicating through representative units and household payments of resource usage, such as water and energy, provides individuals with limited recognition (Bowles, 2008; Jeong et al., 2014). Instead, depicting the energy usage in terms of the amount of deforestation or displaying water usage over a year as a number of swimming pools renders the consumption based on factual knowledge more salient and effectively attracts individuals’ interest and attention (Jain et al., 2013; Novak et al., 2016; Petkov et al., 2011). Numerous studies have further utilized vivid demonstrations through visualizations such as 3D models, photos, and graphs to provoke greater perceptual responses and even personal conservation behaviors (Pahl et al., 2016; Sheppard, 2015).

Immediacy of the consequence

Immediacy is an attribute of visualization that makes longer-term consequences seem nearer-term through accelerating time (Sheppard, 2005). This approach can provide information about the cumulative negative impact on the environment of activities such as water or energy overconsumption as conspicuous signals that individuals rarely observe directly or immediately. Clear, ominous messages of the future conditions dramatically reduce the temporal gap and make the demonstration of water-conserving behaviors more persuasive (Carlson, 2001). Previous researchers have reported that when people perceive future consumption of resources, they are more likely to reconsider future consequences (Hershfield, 2011; Zimbardo & Boyd, 2008).

Personal relevance to the problem

The facilitation of behavioral modification requires disclosure of the personal relevance of the environmental problem (Anderson, 2012; Bator & Cialdini, 2000; Lowe, 2006; Roczen et al., 2014). Numerous studies have provided such relevancy by showing the water or energy consumption feedback from the daily habitual inaction of individuals (Delmas et al., 2013; Novak et al., 2016). These feedback strategies focus on the outcomes of the
specific activities to clarify to individuals that they can influence them (Fischer, 2008; Karlin et al., 2015). Furthermore, when individuals know how to achieve the modifications, they feel more strongly responsible for the problem and are more willing to effect change (Fischer, 2008; Nussbaum et al., 2015).

**Experience in immersive virtual environments**

Experiential learning is a powerful delivery method to enhance cognitive and emotional and behavioral responses to environmental problems. Studies suggest that the information that is presented by experiential methods is more salient and drives higher attention than that provided in the form of description (Ahn et al., 2014; Karthe et al., 2016; Pooley & O’Connor, 2000; Quay et al., 2013; Weber, 2006). For promoting the perception of environmental degradation, personal experience is recommended as an effective approach to reduce temporal and psychological distance (Larson & Redman, 2014; Redman, 2013). Experiencing negative consequences, such as water scarcity, improves people’s attitudes, which are more affectively based (Corral-Verdugo, 2002). Despite the energy of experience-based strategies, real-world experiences of negative environmental consequences are often unavailable or infeasible, so a realistic simulation could provide an alternative solution. Immersive virtual environment technology has been applied in numerous studies to enhance the intensity of experiences.

**Immersion**

IVET provides a first-person view of a three-dimensional virtual environment in a panoramic display so that users can see, hear, and feel the content with higher vividness and realism (Bailey et al., 2015; Blascovich & Bailenson, 2011). The experience creates an illusion of the physical world, and users can feel the virtual events happening to them (Ahn et al., 2016). Due to the sense of presence, the sensory information and scene simulated in a virtual environment increase the user’s attention and engagement, and the experiences tend to be stored as individual memories for the long term (Ahn et al., 2016). The cognitive responses will be recalled and further influence personal attitudes.

**Interactivity**

Interactivity is defined as a process in which users have the freedom to influence the displayed content and receive real-time feedback on their own actions (Sundar, 2004). A virtual environment provides a higher level of interactivity, allowing users to naturally move their bodies, such as walking or moving their heads, to positively control the information received during the experience (Ahn, 2011; Bailey et al., 2016; Sheppard, 2005). Meanwhile, to respond to the users’ actions as in an actual physical environment, the objects in the virtual environment are rendered and controlled with a physics engine. The dynamic experience enhances users’ engagement and enthusiasm and further encourages changes in attitude and behavior.

**Flexibility**

The content of the experience in a virtual environment is boundless. Users can experience scenarios that would be impossible in the real world. For example, a virtual environment can provide sensory experiences of body transfer (e.g., the individual is inside the body of an animal (Ahnet al., 2016), disability (Ahn et al., 2013)), future consequences (e.g., individual see their own aged faces (Oh et al., 2016)) and impossible actions (e.g., being a superhero and saving virtual people (Rosenberg et al., 2013)). The flexibility and extensive control of IVET increases the efficiency of highlighting meaningful messages and making abstract concepts and information concrete. Also, experimenters can easily control the displayed form of the virtual environment and unobtrusively measure every movement made by the users.

Due to the characteristics of IVET, namely, immersion, interactivity, and flexibility, numerous studies suggest that IVET could be a powerful form of persuasive technology for catalyzing personal cognition, attitudes, and behavior changes. IVET can be used to expose users to novel events that can create new beliefs and emotions (Felnhofer et al., 2015; Oh et al., 2016). The effect is caused not only by the levels of interactivity, flexibility and immersion that IVET provides but also by individual differences such as gender and past personal experiences (Ahn et al., 2013; Sandstrom et al., 1998). Experience-learning through IVET has the potential to make the target consciousness easier to achieve while it should be evaluated that personal belief and emotion to the context in
virtual environment. Focusing on enhancing pro-environmental attitudes and behaviors, numerous studies have utilized IVET as persuasive technology to increase awareness of environmental issues.

**Game design**

To examine the hypothesis, we designed a virtual reality water conservation game based on psychological considerations to fill the gap between the individual’s habitual behaviors and the negative consequences for the environment. One study showed that game-based learning results in more positive effects than only using virtual worlds or simulations to foster learning. Game-initiated learning is a persuasive design for motivating students and delivering practical experiences (Tsai et al., 2015). The virtual reality game was designed for two participants. One was the major player in the virtual environment. The other one was the assistant and stayed in the physical environment. The two participants needed to cooperate with each other to complete the tasks in the virtual reality game. Below, we describe the psychological considerations and apparatus used in the virtual reality game.

**Psychological considerations**

The overall considerations in the design of the virtual experience are presented in Figure 1. The content of the virtual reality game corresponded to the key elements of raising consciousness of environmental problems related to water.

![Figure 1. Overall considerations in game design](image)

**Relevance of personal behavior**

Figure 2 shows the virtual bathroom, including a toilet, shower, and faucet, that was built in the game.

![Figure 2. The bathroom in the virtual reality game](image)
In the game, participants were required to exhaust the water in the virtual bathroom for flushing the toilet and taking a shower, after which they received feedback on their water consumption in real time. The design was intended to present a clear connection between specific habitual activities and water consumption to increase the participants’ awareness of how their customary behaviors are related to environmental problems.

**Vivid water consumption**

To make the amount of daily water use more concrete, we utilized 600-milliliter bottles and the consumption of physical energy to represent feedback on water consumption.

*Figure 3. Surroundings in immersive virtual environment, including a central water reservoir, numerous virtual bathrooms, and an oasis*

In the game, as shown in Figure 3, a central water reservoir stored the water resource, and numerous virtual bathrooms were located in the oasis. The water could be used after being drawn to a water-storage tank, as shown in Figure 4, in the virtual bathrooms. To build a water scarcity environment, water could be drawn if the assistant in the game, the other player, repeated the physical motion of pumping to fill the water-storage tank little by little. On the other hand, players in one of the bathrooms were allowed one minute to use a 600-milliliter bottle repeatedly to transfer water from a water storage tank (left side of Figure 5) to specific water tanks for flushing the toilet and taking a one-minute shower. After the tasks, the actual water consumption of each activity was visualized as multiple 600-milliliter bottles (right side of Figure 5).

The vivid feedback, namely, 600-milliliter bottles and the drop in the water level of the central water reservoir, could construct the individual’s knowledge of the actual amount of water used. Moreover, the personal experience of the effort of transferring water created fatigue, which could cause the individual to reflect on the feedback.

*Figure 4. Physical motion of pumping to draw water from the central water reservoir to the bathroom*
Exaggerated feedback (EF)

The exaggerated feedback was intended to convey the concept that an individual may unconsciously cause negative consequences, which is hard to be aware of while using water in daily life. Two types of exaggerated feedback were considered in the virtual reality game, those being an accelerated negative impact on the environment (Ambient EF) and an accelerated negative impact on consumption (Direct EF). Ambient EF (Negative impact on environment feedback) was indicated by shrinking of the oasis due to the drop in the water level of the central water reservoir (Figure 6) Direct EF (negative impact on consumption) was indicated by the rapid consumption and exaggerated drop in the water level. The exaggerated feedback was intended to convey the environmental (ambient EF) and water resource (direct EF) effects of the participant’s consumption to emphasize the importance of saving water.

A 2x2 between-subjects design was implemented to examine the effects of the two types of exaggerated feedback, negative impact on the environment (low vs. high) and negative impact on the resource (low vs. high), on attitude and behavior intention to conserve water, as shown in Figure 7. In the condition with the feedback of negative impact on the environment, the walls around the bathroom were opaque, so the participants were unable to detect the transformation of the oasis into a desert. To prevent participants from unintentionally becoming aware of the change, the virtual scene would become black-and-white if the participants walked out of the bathroom.
Apparatus

In this study, a head-mounted display system with two integrated hand controllers (HTC VIVE) and a motion sensing input device (Microsoft Kinect) were used, and the virtual game was programmed in the Unity and Steam VR toolkit.

Head-mounted display

While wearing the head-mounted display, the participants experienced a three-dimensional environment from a first-person perspective. The head orientation and position of the participants were tracked by two sensors mounted on the ceiling, so the participants could look and walk around the virtual environment with natural physical movements. To intensify the immersion, stereophonic sound effects, including water flowing and realistic sounds of the desert, were also provided through the earphones. Figure 8 depicts the setup of the experiment.
Hand controllers

Figure 9 shows the virtual appearance of the two hand controllers, which allowed users to interact with the virtual environment in real time. The left-hand controller elicited tasks and provided guidance for the participants to finish the virtual experience. After reading and understanding the guidelines, participants could press the button on the controller and continue the game. The right-hand controller was shown as a virtual 600-milliliter bottle, with which participants could transfer the water. During the transfer process, participants received haptic feedback (i.e., vibrations from the controller) while filling the bottle or pouring the water.

![Figure 9. Virtual appearance of two controllers for water delivery and guideline text](image)

(e.g., In the above figure the guideline text says: “Now please ask your partners to press the pump to cause water flow into water-stored box.”)

Motion sensing input device

A motion sensing input device was used to detect the physical motions of the game assistant in real time. The game assistant stood at a distance from the device that allowed the device to track the assistant’s movements. Meanwhile, the assistants could view the perspective of the main player on a 27-inch display. Figure 10 shows the overall experimental setup of the virtual reality game.

![Figure 10. Overall experimental setup: (a) Game assistant stood in front of the Kinect sensor (b) Main player wore the head-mounted display to perceive the virtual environment](image)

Methodology

The experiment was conducted in two phases in order to investigate whether the virtual reality game and exaggerated feedback could decrease the individual's temporal distance from the environmental risk of overconsumption of water to improve the behavior intention to reduce water usage.
Participants

Participants aged 16 to 17 (\( M = 16.30, \ SD = 0.46 \)) years old were recruited from a senior high school in Taipei. The sample (\( N = 165 \)) consisted of 15 female and 162 male students. Students were recruited to participate in an event, Online to Onsite (O2O), for promoting water resource education at the National Taiwan Science Education Center (NTSEC), and all provided informed consent prior to participation. The event lasted for four days, and about 40 students attended each day.

Procedure

The experiment was conducted in two phases. Phase One was conducted on the day on which students participated in the event, and Phase Two was completed approximately one month after the experimental treatment. In Phase One, all participants attending the event on the same day, about 40 students, finished a written pretest to measure their baseline levels of attitude, behavior intention to conserve water, and habits of water usage in a meeting room at the NTSEC. Later, the participants were divided into four groups and assigned to one of four exaggerated feedback conditions in the immersive virtual environment, as follows: without EF (\( N = 37 \)), only ambient EF (\( N = 42 \)), only direct EF (\( N = 45 \)) and ambient direct EF (\( N = 41 \)). Two participants, a major player and an assistant, cooperated in each round of the experiment. As shown in Figure 11, the virtual experience consisted of two parts, a training mission and a main mission. The main mission consisted of three tasks of consuming water.

![Figure 11. Flow chart of the water conservation virtual reality game](image)

The duration of the overall experiment was about 7 to 10 minutes for each participant. At the start of the experiment, the major player was provided with guidance on playing the game to ensure the equal accessibility of the IVET to each participant. The main player stood in the virtual bathroom and viewed an introduction of the game background, as follows: “You are located on an oasis suffering from a scarcity of water resources. We need your energy to deliver the water to complete missions that will make the daily water-consumption facilities work normally.” Then the major player and the assistant completed the training mission, one instance of transferring the water with a 600-milliliter bottle, to ensure that the participants were familiar with the procedure of the virtual reality game. After the game demo, the major player viewed the message, “During the mission, the avatars in other bathrooms will consume water simultaneously.”

Next, the main player stepped into the part of the main mission, which involved depleting their physical energy to afford the habitual water usage. Progression of the virtual simulation is presented in Figure 12. First, the player was instructed as follows: “Please fill the water tank of the toilet in 1 minute and flush the toilet one time”, and then the mission began. At the end of the task, the system examined the water level of the specific water tank, the time left, and the amount of water spilled to report the final game score to the player. The player was also informed of the actual consumption per toilet flush, represented by the number of 600-milliliter bottles.
Then the player was instructed to turn on the faucet to wash their hands by the message, “You have succeeded to flush the toilet. Why don’t you wash your hands?” During the hand-washing task, the player was not asked to turn off the faucet. The player had complete freedom to control the flow of water throughout the game. Finally, the player filled the water tank of the shower in 1 minute and took a one-minute shower, the procedure of which was identical to that of the first task. The final score of each player was recorded and ranked on a scoreboard, and the top five scores were shown at the end of the game. After the virtual experience, participants were then guided to fill out a written questionnaire that measured the level of immersion and collected written feedback on the participants’ feelings and emotions during the game. In Phase Two, one month following Phase 1, the participants completed a written delayed posttest investigating their changes in attitude, behavior intention to conserve water, and habits of water usage in their high school classrooms.

**Measures**

There were five measures in this study: cognition of daily water consumption, attitude toward water usage, self-reported behavior intention to conserve water, extent of tightening of the virtual faucet, and presence in the virtual environment. The first three items were measured before the virtual game (Time 1, pretest) and one month later (Time 2, delayed posttest). The other two items were measured and recall of the virtual experience was tested after the participants had finished the virtual game. The questionnaire was self-developed under the guidance of researchers or according to previous research.

**Cognition of water consumption**

Two multiple-choice questions were used to access personal cognitive correctness of daily water consumption. These questions were developed based on the context in the virtual game, where the water consumption of flushing a toilet and that of taking a one-minute shower were vividly displayed. One question was, “How much water is consumed if a non-water-saving toilet is flushed one time? (1) 1–4 liters (2) 5–8 liters (3) 9–12 liters (4) 13-16 liters.” The second question, on the water consumption of a one-minute shower, was provided in a similar pattern. The responses to the two questions were scored as correct or incorrect. The final value of this measurement was the total number of questions that participants correctly answered.
Attitude toward water usage

Seven 6-point interval scale items (1 = strongly disagree, 6 = strongly agree) were used to gauge participants’ attitudes toward water usage, including responsibility and emotion, based on previous work (Peçanha de Miranda Coelho, Veloso Gouveia, Silva de Souza, Lemos Milfont, Barros, & Nogueira, 2016). Two example items are “I think if I do my best, I can improve or solve the water issue,” and “I feel guilty when I forget to turn off a tap.” The Cronbach’s alpha (reliability) of the seven items was 0.63.

Water saving behavior intention

Six 6-point interval scale items (1 = strongly disagree, 6 = strongly agree) were used to gauge participants’ intentions to reduce water usage, such as by shortening their shower time or actively dealing with water leaks from a toilet or faucet that was not turned off. Two additional items asked participants to report their personal habits of daily water consumption. One item was, “In general, for how long do you shower? (1) less than 5 min (2) 6–10 min (3) 11–15 min (4) 16–20 min (5) more than 21 min.” The other one was related to the frequency of reusing water resources. Three choices, “always,” “sometimes,” and “rarely” were provided. Linear transformation was applied to combine different point scale data. The Cronbach’s alpha of the eight items was 0.75.

Extent of tightening the virtual faucet

In the process of experiencing virtual environment, participants were asked to turn on the faucet for washing hand and then were allowed to continue the game. However, we did not force participants to turn off the faucet that would keep virtual water flowing and wasted. Participants had freedom to deal with the sewage by actively turn off the faucet before the end of the game. The extent of tightening the virtual faucet was measured in percentage and the final number that could be considered as behavior intention in short-term was recorded.

Presence in the virtual environment

Presence was measured with three four 6-point interval scale items (1 = strongly disagree, 6 = strongly agree), which measured participants’ subjective perceptions of the realism of the mediated virtual experience and whether exaggerated feedback intensifying the negative consequences could cause a lower personal perception of presence. The questions were adapted from Ahn et al. (2014); Bailenson et al. (2005) to assess the presence in the virtual environment. For example, the items were “It feel like I visited the water scarcity place.” “I feel the experience of pouring water is real.” The presence was immediately evaluated after participants finish the game. Reliability for the eight items was Cronbach’s alpha = 0.63.

Recall

After finishing the virtual game, the participants were encouraged to recall their cognitive and affective responses in the process of the experience and to write them down on paper. They were provided two open-ended questions: “Which scenario made the greatest impression?” and “How did you feel after the virtual experience?” The written responses were organized according to several key points and are presented in the results section.

Results

We analyzed the data recorded in the virtual reality game and questionnaire results to answer our two research questions. We focused on the overall effect of the virtual reality game and investigated the impact of exaggerated feedback with measures.
**Effect of vivid and personally relevant experience in a virtual environment**

Paired t-test between pretest and posttest was used to determine whether the virtual experience incorporating vivid and personally relevant information had effects on the personal cognitive, affective, and behavioral domains. As Table 1 shows, overall, participants’ changes in cognition ($M_1 = .39 \ (SD = .23)$, $M_2 = .60 \ (SD = .30)$), attitude ($M_1 = 4.16 \ (SD = .83)$, $M_2 = 4.20 \ (SD = .92)$) and behavior intention ($M_1 = 4.07 \ (SD = .66)$, $M_2 = 4.24 \ (SD = .73)$) were positive. The paired t-test also indicated statistically significant differences in the personal cognitive ($t = -3.389 \ (SE = .43), p < .01$) and behavioral ($t = -3.63 \ (SE = .36), p < .01$) domains before and after the experimental treatments. On the other hand, no significant effect ($t = -.886 \ (SE = .49), p > .05$) of the virtual experience on overall personal attitude was revealed.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Time 1</th>
<th>Time 2 (after one month)</th>
<th>$t$-value</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognition to water consumption</td>
<td>$M_1 = .39 \ (SD = .23)$</td>
<td>$M_2 = .60 \ (SD = .30)$</td>
<td>-3.389</td>
<td>.001**</td>
</tr>
<tr>
<td>Attitude to water usage</td>
<td>$4.16 \ (SD = .83)$</td>
<td>$4.20 \ (SD = .92)$</td>
<td>-0.866</td>
<td>.337</td>
</tr>
<tr>
<td>Water saving behavior intention</td>
<td>$4.07 \ (SD = .66)$</td>
<td>$4.24 \ (SD = .73)$</td>
<td>-3.63</td>
<td>.001**</td>
</tr>
</tbody>
</table>

*Note. **p < .01.*

**Effect of exaggerated feedback**

Descriptive statistics of all measures that were analyzed according to different condition resulted from direct EF and ambient EF are presented in Table 2. To investigate the effect of exaggerated feedback, two-way analyses of variance (ANOVA) was run with direct EF and ambient EF as the between-subject variable; and extent of tightening virtual faucet, the change of cognition, attitude and behavior intention between pretest and posttest and personal presence in virtual environment as independent variable. The results from the ANOVA are shown in Table 3.

<table>
<thead>
<tr>
<th>Condition</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>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M_1 (SD)$</td>
<td>$M_2 (SD)$</td>
<td>$M_1 (SD)$</td>
<td>$M_2 (SD)$</td>
<td>$M_1 (SD)$</td>
<td>$M_2 (SD)$</td>
<td>$M_1 (SD)$</td>
<td>$M_2 (SD)$</td>
</tr>
<tr>
<td>Overall not direct EF</td>
<td>.43</td>
<td>.39</td>
<td>.61</td>
<td>4.12</td>
<td>4.20</td>
<td>3.98</td>
<td>4.20</td>
<td>4.54</td>
</tr>
<tr>
<td></td>
<td>(.50)</td>
<td>(.59)</td>
<td>(.69)</td>
<td>(.53)</td>
<td>(.43)</td>
<td>(.68)</td>
<td>(.57)</td>
<td>(.81)</td>
</tr>
<tr>
<td>Overall direct EF</td>
<td>.50</td>
<td>.40</td>
<td>.62</td>
<td>4.19</td>
<td>4.21</td>
<td>4.15</td>
<td>4.27</td>
<td>4.47</td>
</tr>
<tr>
<td></td>
<td>(.50)</td>
<td>(.54)</td>
<td>(.62)</td>
<td>(.61)</td>
<td>(.51)</td>
<td>(.64)</td>
<td>(.60)</td>
<td>(.86)</td>
</tr>
<tr>
<td>Overall not ambient EF</td>
<td>.55</td>
<td>.46</td>
<td>.67</td>
<td>4.28</td>
<td>4.21</td>
<td>4.20</td>
<td>4.25</td>
<td>4.60</td>
</tr>
<tr>
<td></td>
<td>(.50)</td>
<td>(.57)</td>
<td>(.67)</td>
<td>(.55)</td>
<td>(.54)</td>
<td>(.66)</td>
<td>(.63)</td>
<td>(.84)</td>
</tr>
<tr>
<td>Overall ambient EF</td>
<td>.39</td>
<td>.33</td>
<td>.55</td>
<td>4.04</td>
<td>4.19</td>
<td>3.94</td>
<td>4.23</td>
<td>4.41</td>
</tr>
<tr>
<td></td>
<td>(.49)</td>
<td>(.54)</td>
<td>(.63)</td>
<td>(.57)</td>
<td>(.39)</td>
<td>(.64)</td>
<td>(.54)</td>
<td>(.83)</td>
</tr>
<tr>
<td>Not direct, not ambient EF</td>
<td>.49</td>
<td>.54</td>
<td>.65</td>
<td>4.20</td>
<td>4.20</td>
<td>4.17</td>
<td>4.20</td>
<td>4.67</td>
</tr>
<tr>
<td></td>
<td>(.51)</td>
<td>(.65)</td>
<td>(.70)</td>
<td>(.53)</td>
<td>(.51)</td>
<td>(.66)</td>
<td>(.60)</td>
<td>(.79)</td>
</tr>
<tr>
<td>Direct, not ambient EF</td>
<td>.60</td>
<td>.40</td>
<td>.69</td>
<td>4.34</td>
<td>4.22</td>
<td>4.23</td>
<td>4.29</td>
<td>4.55</td>
</tr>
<tr>
<td></td>
<td>(.49)</td>
<td>(.50)</td>
<td>(.55)</td>
<td>(.57)</td>
<td>(.57)</td>
<td>(.66)</td>
<td>(.66)</td>
<td>(.88)</td>
</tr>
<tr>
<td>Not direct, ambient EF</td>
<td>.38</td>
<td>.26</td>
<td>.57</td>
<td>4.05</td>
<td>4.20</td>
<td>3.82</td>
<td>4.20</td>
<td>4.44</td>
</tr>
<tr>
<td></td>
<td>(.49)</td>
<td>(.65)</td>
<td>(.68)</td>
<td>(.53)</td>
<td>(.35)</td>
<td>(.67)</td>
<td>(.56)</td>
<td>(.82)</td>
</tr>
<tr>
<td>Direct, ambient EF</td>
<td>.39</td>
<td>.39</td>
<td>.54</td>
<td>4.03</td>
<td>4.20</td>
<td>4.07</td>
<td>4.26</td>
<td>4.39</td>
</tr>
<tr>
<td></td>
<td>(.49)</td>
<td>(.50)</td>
<td>(.67)</td>
<td>(.61)</td>
<td>(.43)</td>
<td>(.60)</td>
<td>(.52)</td>
<td>(.84)</td>
</tr>
</tbody>
</table>

*Note. 1 = The extent to tighten virtual faucet; 2 = Cognition to water consumption baseline; 3 = Cognition to water consumption post; 4 = Attitude to water usage baseline; 5 = Attitude to water usage post; 6 = Water saving behavior intention baseline; 7 = Water saving behavior intention post; 8 = Presence in the virtual environment.

<table>
<thead>
<tr>
<th>Condition</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F$</td>
<td>$p$</td>
<td>$\eta^2$</td>
<td>$F$</td>
<td>$p$</td>
</tr>
<tr>
<td>A</td>
<td>.59</td>
<td>.44</td>
<td>.004</td>
<td>.005</td>
<td>.946</td>
</tr>
<tr>
<td>B</td>
<td>4.048</td>
<td>.05</td>
<td>.025</td>
<td>.052</td>
<td>.821</td>
</tr>
<tr>
<td>C</td>
<td>.42</td>
<td>.52</td>
<td>.003</td>
<td>.762</td>
<td>.186</td>
</tr>
</tbody>
</table>

198
Note. A = Change with direct EF; B = Change with ambient EF; C = Change with direct & ambient EF; 1 = Tighten virtual faucet; 2 = Personal cognition improvement; 3 = Personal attitude improvement; 4 = Personal behavior intention improvement; 5 = Presence in the virtual environment. *p < .05.

Extent of tightening of the virtual faucet

As Table 2 shows, participants in the only direct EF condition had the highest behavior intention to turn off the virtual faucet, with a mean of the extent of tightening the virtual faucet of .60 (SD = .49). Comparing the overall direct EF condition with the not overall direct EF condition, it is likely that direct EF had a positive impact on the short-term behavior intention to save water. In this study, no significant effects of direct EF (F = .59, p > .05, Eta-squared = .004) or interaction effect (F = .42, p > .05, Eta-squared = .003) were found.

Improvement of cognition, attitude and behavior intention

With regard to the long-term effect of EF, results showed a significant effect of ambient EF on enhancing personal water conservation attitude and behavior intention. The improvement of individual attitude and behavior intention was greater after exposure to ambient EF than to not ambient EF. Table 3 presents statistically significant effects of ambient EF were found for personal attitude improvement (F = 5.034, p < .05, Eta-squared = .03) and personal behavior intention improvement (F = 6.596, p < .05, Eta-squared = .39). No significant interaction effect (p > .05) was found for attitude improvement (F = 4.99, Eta-squared = .003) or behavior intention improvement (F = 1.255, Eta-squared = .008).

Presence in the virtual environment

Presence in the virtual environment was higher without EF (M = 4.67, SD = .79) than with EF, including the only direct EF condition (M = 4.55, SD = .88), only ambient EF condition (M = 4.44, SD = .82), and direct ambient EF condition (M = 4.39, SD = .84) as Table 2 shows. However, the effect of EF on presence was not significant. The results indicated that EF did not dramatically decrease the personal feeling that the virtual experience was genuine.

Immediate recall of virtual experience

Recall was collected from two open-ended questions immediately following the virtual reality game. The responses mainly fell into three categories: engagement, realism, and awareness. Engagement-related responses, comprising about 52% of the response, was focused on the entertainment and accessibility of the virtual game. Examples of responses are, “It was really interesting!” “The task in the game was a little difficult,” and “The experience was very attractive to me.” About 38% of the feedback (35% positive, 3% negative) depicted the realism of the virtual game, such as “The experience was more realistic than expected,” “It was just like being located in an environment suffering from water scarcity,” and “Pouring water in the virtual environment was realistic.” Such responses can be regarded as support for evaluating the personal presence. About 10% of the responses was classified as awareness. These statements were expressions of personal emotion about the exaggerated virtual experience and cognition regarding conservation of water issues, such as “The experience gave me a quite shock” and “Although the game was really fun, it informed me of the importance of saving water.”

Discussion

Reason for positive effects of the virtual reality game

The purpose of the present study was to create virtual experiences that exposed individuals to vivid information in hopes of increasing the behavior intention to conserve water. Thus we evaluated the effects of a compelling immersive virtual environment on behavioral changes, echoing earlier studies on climate change communication by demonstrating that interactive immersion in a virtual environment was more persuasive and would likely motivate behavioral intent (Sheppard, 2005). We compared personal levels of cognitive, affective, and behavioral responses to water conservation issues one month after the mediated virtual experience with outcomes from the
The results indicated that the experience of depleting one’s physical energy to afford daily water usage in the virtual environment may successfully convey to individuals a strongly vivid message. The vividness caused by making abstract water consumption concrete with 600-milliliter bottles and even the direct experience of water transfer in the game facilitated students’ clear comprehension of the actual amount of water consumed every day. Moreover, regarding the significant effect on increasing behavior intention, it is possible that the familiar scene, using water resources in a bathroom, provided the specific cause of excessive consumption and then raised awareness of personal relevance to the water conservation issue. The open feedback, such as “From an objective perspective, I am surprised that daily water consumption is so high” and “We always consume a lot of water unconsciously,” also provided similar results. Therefore, seeing vivid causes and clear effects, the students were more willing to engage in water conservation behaviors. Similar to earlier studies summarized in Larson and Redman (2014), the current study launched programs and conducted experiments attempting to transform behaviors toward conservation and sustainability. The central idea was to raise awareness that would lead to corresponding responses and reflections by presenting contextual issues regarding environmental protection.

Previous studies (e.g., Dillard & Main, 2013; Sheppard, 2012) have utilized the “vivid effect,” referred to as exaggerated feedback (EF) in our work, to significantly accelerate individual understanding of the actual consumption and make messages more persuasive. Also, 3D models, photos, and graphs have been developed to provoke greater perceptual responses and even personal conservation behaviors (Pahl et al., 2016; Sheppard, 2015). We also took advantage of the fact that “exaggeration is a malleable concept.” Specifically, the “exaggerated feedback” was proposed with the hypothesis that it had the “power” to trigger many aspects of human beings’ cognition and behavioral responses. On the other hand, the overall attitude toward water usage was not significantly improved. The lower effect may have been due to the change in personal attitude not being largely determined by vivid and personally-related information, which will be discussed in the next subsection.

**Reason for positive effects of exaggerated feedback**

Two kinds of exaggerated feedback, direct EF and ambient EF, were controlled in the experimental treatments. In the following, to answer Q2, we explore the effects of exaggerated feedback during the process of experiencing a virtual environment (at Time 1) and one month later (at Time 2) according to corresponding independent measures.

**Immediate effects**

The immediate effects of exaggerated feedback were investigated with the personal intention to turn off the virtual faucet. Although participants exposed to direct EF tended to tighten the faucet, the outcomes had no significant differences from the no direct EF condition. A reasonable explanation is that there was no pretest treatment to evaluate the baseline of individual water saving behavior in the virtual environment. The baselines of attitude and behavior intention in the separated groups, with and without ambient EF, were validated to have a significant difference ($p < .05$), as was the difference ($p < .05$) in attitude baseline in the direct EF and no direct EF conditions. The results showed that, without controlling for students’ prior behavior in the virtual immersive environment, the outcomes of this experimental treatment were changeable.

**Delayed effects**

The results of this study show that ambient EF and accelerated environmental degradation directly caused by personal behavior can stimulate a pro-water-conservation attitude and behavior intention. The experience with ambient EF somehow had a lower level of realism than that of the no EF condition, but it did not dramatically increase personal disbelief in the water conservation issue. One deduction might be that the open feedback about realism was nearly all positive in the exaggerated condition. With enough presence in the experience, students’ affective responses, including their guilt over wasting water and sense of responsibility for water conservation, were prone to be evoked by emphasis on the change in the virtual ambient environment, which facilitated the intention to modify their behavior. The open feedback from students also generated the expected reflections, such as “The situation in the virtual game could happen in the near future, so we should be more engaged in water conservation,” and “Quite shocking! It made me understand the inconvenience in a water shortage area.”
Unexpectedly, no effects of direct EF on cognition, attitude, and behavior intention were found. It is likely that direct EF, achieved through exaggerating the water consumption while the participants filled the 600-milliliter bottles, did not make the immediacy of the negative consequences salient. Compared with ambient EF utilizing the wrap-around display of virtual reality, direct EF was less intuitive for participants and did not direct their attention to the negative consequences of their behavior. To provide access to mental processes through direct EF, related basic knowledge should be provided as a baseline to heighten the contrast with exaggerated outcomes.

We specifically focused on how to implement exaggeration in designing virtual reality games to influence participants’ attitude and change behaviors in facing issues in environmental education. Unlike the tools and facilities used in previous studies, the new science and technology of VR was used in this study to provide a more vivid visual presentation with the goal of raising the individual’s personal relevance and immediacy in responding to water resource issues. Previous studies have shown that individual behaviors may directly cause negative consequences for many aspects of water resources, indicated as direct EF in the current study. Here we further elaborate the effect by examining the impact on the environment as a whole, indicated as ambient EF. Parts of our experiments had relatively small significant outcomes, possibly due to the nature of the virtual reality games. Participants inevitably were resistant to some extent, perhaps feeling that these exaggerations were simply designed to force them (or made it easy for them) to choose from one or another situation. However, the unique contribution of our study is that it integrated and applied the concept of exaggeration in game design in the environmental education field.

**Conclusion**

In this study, to bridge the perceptual gap between behavior and actual consequences in water conservation education, we developed an immersive virtual water conservation game as persuasive technology for experiential learning. The virtual reality game, which involved using one’s physical energy to deliver water for use in a bathroom, provided individuals access to vivid and personally relevant experiences of water consumption. The virtual experiences allowed participants to better process their understanding of actual water consumption and increase their engagement in water conservation.

Furthermore, focusing on the temporal distance of the water conservation risk, exaggerated feedback in the embodied environment was suggested in this study, and it was applied to emphasize that individual behaviors of daily water usage may directly cause negative consequences on water resources (direct EF) and the environment (ambient EF). The evidence revealed that participants exposed to the ambient EF with the immediacy of the degradation of the virtual environment tended to have higher levels of affective response and intention to modify their behavior for conserving water, whereas the direct EF did not provoke similar responses. We found that ambient EF, achieved by gradually degrading the virtual surroundings while the water was being exhausted, was more intuitive and salient than direct EF, thus leading to success in evoking emotions and raising personal awareness of the water conservation issue.

In summary, this study contributes a novel experiential intervention in water conservation education. The intervention utilized the characteristics of IVET, namely, immersion, interactivity, and flexibility, and the self-designed context in the game achieved the key elements: consciousness of the water conservation issue, vivid effects, personally-related information, and immediacy of consequences. Furthermore, the use of exaggerated feedback in a virtual environment appears to be an effective strategy for promoting intention to actively engage in pro-environmental activities.

**Limitation and future work**

A lingering question is the connection between each response to virtual experience, which remains unclear. The current study found positive impacts of the virtual experience and exaggerated feedback on cognition, attitude, and behavior intention, but the psychological model, such as cognitive effects on attitude or behavior intention, and the extent of responsibility and emotion aroused, should be investigated in detail.

Regarding the experienced context in the virtual environment, the current study mainly assessed participants to loss frame that stimulated individual guilt and anxiety to increase concern about environmental issues. Although the virtual reality in this study may have made it clear to the participants what daily behaviors cause excessive water consumption, which in turn showed individuals how to avoid the negative consequences, no elements for
raising positive emotion were included in the intervention. The effectiveness of a gain frame with exaggerated feedback should be investigated and compared with the negative one in the future.

Limitations of the experiment design were noted in two aspects: the sample and the frequency of intervention. The sample, consisting of 90% male senior high school students, limits the generalizability of the results. Gender can cause differences in the access to virtual reality technology. Indeed, measures of personal daily behaviors in the physical environment were lacking in this study. To yield further insight into the effectiveness of the water conservation virtual reality game, future research should recruit a wider range of participants from different populations. In addition, the outcomes of repeated interventions and measurements of the environmental behavior have to be considered.

References


Gender-Related Differences in Collaborative Learning in a 3D Virtual Reality Environment by Elementary School Students

Yi-Lien Yeh, Yu-Ju Lan* and Yen-Ting R. Lin

The Department of Chinese as a Second Language, National Taiwan Normal University, Taiwan // sasa0426@gmail.com // yujulan@gmail.com // roylin1003@gmail.com

*Corresponding author

ABSTRACT

This study aimed to understand how children collaborated to create their own stories in a 3D virtual reality (VR) environment and to explore whether there were any gender-related differences in this learning setting. A total of 65 students (38 females and 27 males) of grades 4-5 participated in this study. A qualitative methodology was adopted in this research to answer three research questions about how the participants work collaboratively, whether a gender difference exists in the process of 3D construction and how they think about learning in a 3D virtual environment. Based on the behavior patterns of these participants in a 3D world, the research findings revealed two types of collaboration: a team either with or a team without a strong leader. In addition, gender-related differences in this learning environment were also depicted in this study.

Keywords

3D virtual environment, Collaborative learning, Gender-related differences

Introduction

Traditional teaching and learning approaches and learning theories have been challenged for decades. It has been claimed that game-based learning suits learners of the new generation (Hsieh, Lin, & Hou, 2015; Rankin & Edwards, 2017; Reinders & Wattana, 2014; Watson & Yang, 2016). The advantage of game-based learning has been recognized (Charsky, 2010; Jarvin, 2015), which is that learning and entertainment occur at the same time. Additionally, several researchers have been using technology to enhance learning performance or to increase learning motivation through collaborative learning over the last two decades (Ding, Bosker, & Harsskamp, 2011; Ludvigsen, 2016; Oliveira, Tinoca, & Pereira, 2011; Pellas, Kazanidis, Konstantinou, & Georgiou, 2016; Vogel, Vogel, & Cannon-Bowers, 2006; Warschauer, 1997). It was found that gender difference existed in the behaviors of the students who played a mini-game in school (Hsieh et al., 2015) and the behaviors of those who played lego blocking creations (Kato & Morita, 2009). Apart from the setting of game-learning, sex difference was also found in the art creations (Ahlawat & Budhiraja, 2016; Huynh, Doherty, & Sharpe, 2010). The nature differences between the two genders had an impact on their teacher-student relationship and academic performance (Hajovsky, Mason, & McCune, 2017), which teachers should be aware of. Although gender differences in the computer-supported collaborative learning (CSCL) and 3D virtual learning settings have been investigated, the results are mostly based on the analysis of verbal contents (Prinsen, Volman, & Terwel, 2007) and the results mostly include a framework for evaluating collaborative learning interactions (Daradoumis, Martínez-Monés, & Xhafa, 2006). The importance of gender sensitivity toward educational games has been strongly recognized by Robertson (2012). Furthermore, gender differences in game preferences and game playing habits have been explored (Bourgonjon, Valcke, Soetaert, & Schellens, 2010; Hayes, 2008). The findings of the previous research on CSCL and 3D VR worlds, however, did not seem to have depicted gender differences particularly in learner collaboration for the creation of their own works in a 3D VR world. This issue may impact how teachers develop their pedagogy. Therefore, the purpose of this research is to explore, based on video observation and their thoughts of this learning experience, the gender differences in the context where children created their own films in a 3D VR world.

Literature review

Siemens (2005a) claimed that learning theories should be reevaluated because learning nowadays has been greatly influenced by technology. Learners may nowadays be more likely to rely on computers to store and retrieve information. The focus of learning, rather than simply mugging up knowledge, should be placed on locating appropriate resources and wisely determining whether the information is valuable to be processed. Therefore, the connectivism theory (Siemens, 2005b) was proposed to explain several new learning phenomena. This learning theory states that learning could be defined as an ability to access online resources and to interact with others. The connectivism theory concerned a variety of learning approaches, including “through communities of practice, personal networks, and through completion of work-related tasks” (Siemens, 2005a, p.
1). This theory was similar to situated learning (Lave & Wenger, 1991), which takes knowledge as constructed through the participation in the social practices of communities. Felix (2002) provided a more educational explanation for situated learning: “Skills and knowledge are best acquired within realistic contexts and authentic settings, where students are engaged in experiential learning tasks” (p. 3). These learning theories suggest that researchers should explore diverse concepts of learning, such as social interactions in a learning process and also the complexity of learning contexts.

Although the connectivism theory pointed out the impact of new technology on learning, further significant learning theories, including constructivism, social constructivism, and situated learning, are still needed to understand learning in a 3D virtual reality environment. These learning theories have been remarked by some studies (Glancy & Moore, 2013; Pellas et al., 2016). According to Mikropoulos and Natsis (2011), constructivism is the current theoretical model most of the computer-supported educational setting is based on. Pellas and his colleagues (2016) explained that “learning is a subjective and internal process of building meanings and it is considered as the result of organization and adaption of new information into the existing knowledge that students already have to enhance their experiences” (p. 4). This definition is similar to the one generated from social constructivism (Vygotsky, 1978). These theories suggest that learning is an individual and internal changing process, by the stimulus of interaction, either with people or environments. Given that the collaborative learning approach has potential for encouraging learners to express their own ideas and to critique others’ opinions during the process of collaborative problem-solving, collaborative learning has been supported by social constructivism as a research focus for two decades (Caseo, 1992; Ludvigsen, 2016; Stahl, Koschmann, & Suthers, 2006). According to Stahl and his colleagues (2006), the history of group learning could be traced back to 1960. Ludvigsen (2016) approached this concept from a cognitive level. Based on his literature review, he stated that shared understanding and mutual modeling were crucial to building common ground for action, which is important to collaborative learning. Additionally, while investigating the ideas about collaborative learning, more should be explored from the perspectives of personal positioning in the group, in-group interaction to find out whether the shared value exists among members.

Interaction in an authentic context, in addition to the collaboration among peers described above, is also a key component of social constructivism. 3D virtual reality has been recognized as a potential technology capable of providing learners with authentic contexts and immersive experience for them to carry out collaborative tasks (Lan, Kan, Sung, & Chang, 2016). Additionally, from the practical perspective, the advantages of learning in a 3D learning environment have been identified by several researchers (Chang, Gütl, Kopeinik, & Williams, 2009; Hansen, 2008). These benefits include encouraging students to actively interact with learning contexts, reducing the anxiety of learning, and providing a relatively authentic learning experience. Theoretically, a 3D virtual reality environment is a friendly platform for researchers to explore the concepts of collaborative learning (Andreas, Tsiasios, Terzidou, & Pomportsis, 2010; Cen, Ruta, Powell, Hirsch, & Ng, 2016; Lee, 2009; Roschelle & Teasley, 1995; Sutcliffe & Alryes, 2012). With the technology of text- and voice-based chat and online voice recorder, researchers can collect data of social interactions easier. However, Lee (2009) mentioned that some of those studies which have addressed collaborative learning simply presented the perspectives of the users rather than the concepts of collaborative learning. For example, Chang, Gütl, Kopeinik, and Williams’s research (2009) aimed to evaluate collaborative learning settings in 3D VR worlds. The focus of the research findings is mainly related to the opinions of the participating students. Furthermore, although the users can freely access this virtual world without the limitation of time and space so they could “meet” people from different countries, Jarmon (2009) and Whitelock, Romano, Jelfs, and Brna (2000) were concerned that the cognitive load of mastering this new technology may distract students from learning the target subject.

Regarding the possible challenge for learners as they used 3D virtual worlds as a learning platform, Dickey (2005) found that the identity of avatars may help students reduce their learning anxiety because the learners may be too engaged in role-playing to associate their past unpleasant learning experiences with new ones. This finding was consistent with Lan’s research finding (Lan & Lin, 2016). Finally, this platform allows students to put their ideas into practice, which could strongly motivate them and allow them to learn from errors (Beard, Wilson, Morra, & Keelan, 2009). Furthermore, practice in a 3D virtual environment could help learners to transfer their experience in the virtual environment into the real life (Hansen, 2008).

In contrast with the aforementioned positive effects of 3D virtual reality on collaborative learning, the issue of gender differences in the process of collaborative learning in 3D virtual reality seems to have not obtained consistent results. Some researchers believed that gender differences existed in personality traits (Schultheiss & Brunstein, 2001) and emotional responses (Osborne, 2001), while some thought that sexual differences were shaped by social and cultural influences (Putrevu, 2001). There were biological and developmental differences between hemispheres and lateralization of a human brain, which affected both spatial and verbal skills (Geary,
1996). Boys may perform better in spatial tasks, while girls may have greater verbal skills (González-Gómez, Guardiola, Martín Rodríguez, & Montero Alonso, 2012). Li and Kirkup (2007) found that male students in China and the UK showed higher self-confidence in computer skills. Cuadrado-García and Ruiz-Molina (2010) pointed out that women used computers for social purposes and paid more attention to communicative activities, while men considered computers as machines and focused more on operating systems and functions. Additionally, boys spent more time playing computer games, which may extend boys’ technical knowledge (Adamus, Kerres, & Getto, 2009). Whether the aforementioned difference between male and female computer learners can be also identified when boys and girls collaboratively construct their 3D virtual contexts as well as create their own stories is still an issue insufficiently investigated. Based on the aims of this research mentioned in the introduction and the literature review, the research questions were proposed as follows:

- How did students work collaboratively in a 3D VR environment?
- Were there any gender-related differences in collaborative creations in this learning setting?
- What were their perceptions of learning through collaboration in a 3D VR world?

### Method

#### Participants

Sixty-five students who studied in grades 4-5 in an elementary school in Taipei, Taiwan participated in this study. They were 38 females and 27 males. The participants were divided by genders into groups of 4-6 depending on the class size, forming 5 groups of boys and 5 groups of girls. All the participants had been learning basic computer skills such as typing, drawing and surfing the web for one or two years. Thirty-eight per cent of the participants had prior experiences in 3D construction and role-playing in a 3D virtual environment, such as *Minecraft*.

#### Research design

A qualitative research was adopted in the current study. In order to explore the potential gender differences during the process of 3D construction, all the students’ group creations, including 3D contexts, story scripts, and role-play screen videos, were collected. How they collaboratively worked with group members will be described in the section of Procedure. All the students were interviewed to understand their perceptions of using Build and Show (B&S), which will be described in the section of Instruments. All the collected data were then analyzed by two researchers to confirm the concerned gender differences.

#### Instruments

*Build and Show (B&S)*

The platform used in this study, Build and Show (B&S), was developed by the authors. It was programmed by OpenSim. B&S is user-friendly. The users, hence, were not required to have prior knowledge in programming or 3D modeling for mastering the functions of B&S. Additionally, B&S can be installed in most of the computer, which means that there is almost no extra cost to set it up. Challenge of obtaining the technique was an obstacle to hinder teachers’ willingness to incorporate games into their teaching (Watson & Yang, 2016). B&S can be a possible solution to this concern. As shown in Figure 1, students had their own avatars and created their story scenes by selecting “build.” Next, they could choose what they wanted to build and adjust the locations of the objects. When they were collaboratively building up their scenes, they could communicate through either voice- or text-based chat tools. In this study, establishing an airport was the main task for the class. This 3D setting involved five scenes, namely a safety check area, a fast food shop, a duty-free shop, a customs checking area, and a toilet. Each group of students was assigned to one scene of the airport, and then they had to make a real-life story using their assigned area as the setting. One student of each group was responsible for screen recording of the whole process of their group discussion and their scene establishment. The data from this platform were collected in the form of videos and were analyzed to answer research questions one and two.
The interview questionnaire

The researchers adopted the usability questionnaire (Lewis, 1992) and then adjusted the questions according to the current research context. The aim of this questionnaire was to understand participants’ perceptions, as well as their prior experience of playing VR games. In order to collect students’ active responses, the participants were interviewed in groups. Two advantages of this interview method were considered here. Firstly, the researchers could explore students’ perceptions in details to overcome the limitations of short answers in a paper-pencil questionnaire. Secondly, the participants might be more comfortable with the way of speaking in groups where peers may provide emotional support. Furthermore, in order to probe deeply into individual’s ideas about this learning experience, the participating students were also asked to write down their opinions in the final week of this study. The written feedback and the results of the interviews were interwoven to answer the third research question.

Procedure

This research lasted for 5 weeks from April to May 2015. Figure 3 shows the timeline and the weekly learning goals. The students worked on the platform, B&S, in the first four weeks, and were interviewed and completed the questionnaire in the fifth week.
According to Creswell (2012), “qualitative research collects data in a natural setting sensitive to the people and place under study” (p. 37). Furthermore, the current research aimed to understand the gender differences in collaborative learning and to explore students’ learning experience and their ideas about this learning activity. A qualitative methodology was therefore adopted in this study to reach the research goals. More specifically, the interpretivism was the philosophical position underpinning the data analysis. Considering the age of the participating students, the authors mainly adopted the narrative analysis, which paid more attention to what people responded, rather than how they responded (Riessman, 2005). Having this research assumption in mind, the researchers carefully chose the relevant articles in the literature review in order to clarify the definitions of concepts involved in this study. The comparative matrix for the analysis was based on the research addressing collaborative learning in three main topics: characteristics of communication (Oliveira et al., 2011), process-solution analysis (Bravo, Redondo, Verdejo, & Ortega, 2008), and group functioning (Daradoumis et al., 2006).

Three analysis stages were involved in this study. The first round of analysis was to analyze screen videos whose total length was around 6 hours to reveal the significance within the raw data. Those screen videos lasting less than ten seconds and were mistakenly recorded were removed. The open-coding scheme, field notes, and the transcripts of student’s last week of discussion were generated in this stage. The initial coding scheme was emerged in this stage based on the three broader concepts: communication, process-solution, and group functioning. In the following stage, the two coders watched and evaluated the videos based on the coding scheme and completed their own coding separately. Next, the two coders and the researchers discussed whether the codes should be deleted or added when there were discrepancies in the coding result. The differences in the coding results were mainly caused by different interpretations of certain concepts or judgments of how long a behavior lasted. The kappa coefficient of agreement of the coding scheme obtained from the two coders was .86 ($p < .01$), indicating that the inter-rater agreement reaches a substantial level. As for the analysis of the interview questionnaire, the data were analyzed in two ways. First, with regard to their previous experience which was categorized by multiple choices, the results were presented in a numerical way. Second, the oral data from the interview were transcribed and coded. The procedure of the coding of this data was similar to the how the visual data were coded. The kappa coefficient of agreement of the coding scheme of this questionnaire was .88 ($p < .01$). The final stage was to interpret and explain the findings obtained from the first two analysis stages.

Table 1 lists the relations between the final analysis and the research questions and also displays the different types of research evidence used to answer the research questions. The codes emerged from the raw data and then were aggregated into meaningful units. Based on this coding scheme, we could identify the real situation of...
students’ behaviors while exploring and collaboration in the 3D VR environment. In order to explore the gender-related differences, the frequencies of the codes in the genders were analyzed. One limitation in data analysis should be noted here. The analyzed screen videos were recorded by one student of each group, so the observation of avatars in the screen was mainly limited to a certain angle in the 3D setting. However, with the help of the soundtrack, the verbal interactions could be researched. Additionally, the numbers of codes were not significantly great based on the nearly-6-hour video observation because the occurrences of certain behaviors lasted a long time. For example, the discussion of the female group (toilet) about the storyline lasted for almost 13 minutes.

<table>
<thead>
<tr>
<th>RQ1</th>
<th>Analysis of the behaviors of avatars</th>
<th>RQ2</th>
<th>The frequency of codes of two genders</th>
<th>RQ3</th>
<th>Analysis of participants’ group discussion</th>
<th>Researchers’ field notes</th>
<th>Students’ written feedback and verbal answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Codes</td>
<td>Codes</td>
<td>Codes</td>
<td>The frequency of codes of two genders</td>
<td>Reference</td>
<td>Codes and their frequency</td>
<td>Reference</td>
<td>Reference</td>
</tr>
</tbody>
</table>

**Results**

Based on the process-solution analysis (Bravo, Redondo, Verdejo, & Ortega, 2008), the results of students’ collaboration were presented in two main domains: the process of their collaboration and the characteristics of their final co-creations. Moreover, in order to answer the research questions effectively, the results will be presented in three aspects. The first is the question related to how the students work collaboratively in the B & S. The participating students mainly worked together by controlling their own avatars and communicated by their headsets and microphones, rather than face to face. Therefore, the results of the first question were based on the screen videos to present the analysis of the avatars’ behaviors and the content of their verbal expressions. The second question concerned the gender differences of the collaborative creations which include the students’ skills learned in this experience and also their final video works in this study. Finally, the feedback from the participants was presented.

**The gender differences in avatars’ behaviors and the levels of verbal expressions**

An example of the structure of the analysis is shown in Table 2. In the following paragraphs, the gender differences will be explored according to this coding scheme. All participating students were very excited when they logged in to B&S. A few classes needed stronger class management to assure that the students clearly understand their learning task. The most striking behavior observed in the first lesson was that almost all the students attempted to fly and explore the VR world. This could be seen frequently in almost all the screen videos, and “tell me how to fly” was the most frequently asked question during this program. The first gender difference lay in the way the students explored the 3D VR world by flying. The boys in this study tended to spend longer time exploring the environment. Based on the analysis of the video, the area where the boys explored seemed to be wider and higher than that explored by the girls. Additionally, the boys mastered the techniques of flying their avatars and changing the perspectives of their avatars quicker than the girls did, while the girls tended to be easily confused between their avatars’ perspectives and their own perspectives.

Verbal data included two levels, namely individuals and interactions. More verbal data were collected from the female groups than from the male groups. There were two explanations for this. One was related to the technical error of the recording software. Five soundtracks of twenty-one screen videos from male groups were missing. This may influence the coding result in terms of verbal expression, while the behaviors of avatars were still recorded. However, even so, according to the rest of the sixteen videos from the male groups, the gender-related differences were still clearly observed. First, the boys tended to verbally check the condition of the instruments, such as screens or headsets. After checking, they spent more time on creating and controlling their avatars and seldom on verbally discussing anything with the team members. On the other hand, the girls raised more murmurs of discontent, including the noise from their headsets, or wondered what they should do. This result could be seen in the categories of checking and complaining in Table 2 (male: female = 9:20). Almost all the verbalization of actions of an avatar, for example, “I am falling,” was found in the female groups. In addition, the frequency of interaction between the members was found to be higher in the female groups (male: female = 6:25). The girls showed more instances of asking questions, communication, squabbles and impolite language.
In terms of collaboration, the girls demonstrated a stronger sense of collaboration. It was more often observed in the female groups that they were looking for members and also reminding group members of their roles in the learning task. Additionally, when most avatars of each group gathered in the assigned working area, the girls tended to verbally discuss how they should create their own story collaboratively (4 of 5 female groups), while the boys demonstrated a higher tendency towards finer scene creations and object adjustment in the working area individually. In other words, girls seemed to have participated via more talks, and boys via more actions. However, off-task behaviors could be seen in both genders.

Two types of collaboration were identified in this study: the groups with a strong leader and the groups without a strong leader but peer discussion. In the former type, the group with a strong leader, the occurrence of conversation, discussion, and the quarrel was much less frequent than that of in the latter type, the group with peer discussion. Most members of the group with a leader followed the requests of their leaders without much questioning. On the other hand, the peer discussion groups demonstrated different characteristics, involving more off-task chats and focused discussion. One example from the focused discussion group (working in the virtual toilet area) is shown below.

S1: How to record our performance?
S2: We should practice before recording!
S3: Shall we start the rehearsal?
S1: That’s right! Let’s practice. Raise your hand if you agree!
S2: Ok, let’s practice. Don’t video-record yet! Ready? 321, go. I want to be the narrator!
S3: Stop! Are you sure you want to continue typing? (Some members argued by text-chat about the best timing to start recording.)
S2: It’s the director’s responsibility to control the recording software!
S1: Be quiet! Shall we restart?
[They took turn narrating the story of Little Red Riding Hood. All their avatars were in a toilet.]
S3: Is it appropriate to film this story in a toilet?
S2: But we will need to reassign the roles (if we change our story)
S1: Stop arguing. I know it is difficult to create a story in a toilet.
S2: How is it possible to film a story in a toilet? Teacher! (Shouted loudly)
[Some students from the other groups provided their ideas for this group]

From the excerpt, each member in this group had an equal amount of responsibility and they created a friendly atmosphere to allow themselves to express their own ideas. They created a story about sexual harassment, which was an appropriate storyline for their assigned area.

The characteristics of collaborative creations between female students and male students

Two categories were considered as the learning outcome in this section: the mastery of computer techniques and the short films created by the students. The boys could smoothly adjust between avatar’s perspectives, camera angles and their own perspectives in a shorter time. Additionally, the following computer skill was only demonstrated by boys: checking all the details of their assigned areas, such as to what extent could they change the sets of buildings and objects. Some of the boys attempted to hide or delete the objects. For example, a boy flew around the whole airport and landed on a building outside their assigned area. The boy smoothly adjusted...
between the perspective of the avatar and the user himself. His avatar looked down and then the boy tried to hide the building. The building disappeared and the avatar fell on the ground. In general, the boys tended to test as many functions in a 3D world as they could use, while the girls tended to work collaboratively mainly by verbal communications rather than by creations of new objects for their scenes.

Additionally, it could be seen in almost all the screen videos that one avatar in a different color from the color of the nearby avatars appeared among the small groups, and it was usually in blue or green (the colors assigned to boys). Interestingly, the boys did not mind that there was an avatar from the other groups in their working areas, while an avatar from the other groups would trigger females’ strong actions. The female students would chase the avatar away verbally and push off the avatar, seen as an intruder, physically in the virtual environment. On the contrary, the skill observed in both genders was to simultaneously communicate by text- and verbal-chat. However, the purpose of using the two communication tools was different. The boys only used the text-chat for fun to show naughty language, but the girls created another line of conversation which was different to their verbal communication. For example, when the female group (toilet) verbally rehearsed for their story through their headsets, some of the team members discussed whether they needed to record the process of their rehearsal by text-chat.

As for students’ co-construction screen videos, in both genders, 4 of 5 groups completed their tasks in time. The students’ artifacts stretched their imagination by involving a great sense of humor in their stories. One of the male groups (assigned area: the hamburger store) did not simply demonstrate the transaction process but also embroidered the story with the latest news which was about cooking oil made from recycled restaurant waste. One female group (assigned area: customs control) developed their story based on the reality, including checking the passport and asking for some details of the journey. However, the customs officer recognized her long-lost daughter and son while inspecting them, which created a dramatic ending. Interestingly, one boy groups held a very different mindset of creation in this VR environment. These boys paid more attention to the details of how they arranged their objects, such as adjusting the angles of a computer on the x-ray machine. In their final work, they spent a long time having their avatars queue in front of the machine without saying a word, like a show of an army march. These research data demonstrated the diversity of students’ ideas towards creations in a 3D VR world.

**Students’ perceptions of this learning experience**

Almost all the participating students considered this learning experience special and interesting: 92% of the male students and 98% of the female students expressed that they enjoyed creating in this platform. That was their first experience of exploring in a 3D VR world in school and they enjoyed the freedom to work collaboratively with little intervention by teachers. Gender differences also exist in their feedback on this learning experience. According to the group interview, 99% of the female participants believed that this experience enhanced their creativity, while 74% of the male students thought so. Based on their written feedback, some of the male students showed their motivation to learn English in order to use this platform more efficiently, and almost all male students were happy to learn more computer skills. Only one male student mentioned that he enjoyed working with others, while 8 female students evaluated this experience as a good opportunity for learning to work as a team and also recognized the importance of collaboration. One example of these responses was “This program encouraged us to develop our sense of responsibility and creativity. I believe that both are important for our future. The sense of responsibility is the key to building up trust in a relationship.” It was revealed that they still grasped certain key aspects of collaboration in the four weeks. In addition, as more girls encountered technical difficulties in this setting, these girls obtained the sense of achievement by mastering the skills. One of these female students even underwent a change in her learning attitude. She wrote, “When we were asked to complete some tasks, I sometimes found the tasks difficult. I knew that I have to conquer these difficulties. It was all about my attitude. If I keep trying, it means that I am successful.” In brief, the focus of the internal process of building meaning of this learning experience was also different between genders. The male students focused on what they could do, while female students reflected on what they could learn from this experience and expressed their personal feelings.
Discussion and conclusion

Gender-related differences in the social interactions in the process of co-construction in a 3D VR world

Geary (1996) found that boys were more interested in spatial exploration and putting their own ideas into practice, while girls were more comfortable with verbal skills, which was also revealed in the coding results in Table 2. In this study, boys tended to be more flexible with the detailed changes made by the other members. According to the video observation, male students were more likely to negotiate by directly changing the objects, while females preferred to reach a common ground through verbal interactions, including asking, chatting, and quarreling. As for quarreling, the result of this study contradicted the statement of Prinsen, Volman, and Tercel (2007), who believed that tension and abusive language were commonly observed in male groups. However, in this study, it was much more often found in the female groups than in the males’ (see the code of squabbles and impolite attitudes in Table 2). Interestingly, quarreling or impolite language seemed to be an acceptable communicative style in these female groups. These girls were not obsessive about the form of language or their naughty attitude. Therefore, they still could continue discussing their ideas about the creations. It should be noted here that we need more research on whether this phenomenon only appears in the process of their co-construction or this style is a part of their daily talking habits. This finding implies that teachers can set a clear learning target without much intervention during their collaboration and also allow students to negotiate in their own ways, rather than limiting their speaking manners at that moment. Moreover, as the previous and current studies have recognized the gender differences in the collaboration both in a real world and in a 3D VR world, teachers should provide different types of learning tasks to encourage students to develop multiple skills, especially in a 3D VR world. For example, one type of learning task should be designed to guide male students to verbally evaluate their actions after their co-construction, thereby developing their communication skills (Oliveira et al., 2011), while another type of learning task should be provided to encourage female students to develop their spatial exploration.

Gender-related differences in the style of collaboration

Although gender-related differences in CSCL have been explored by Prinsen and his colleagues (2007), the focus of their research was more concerned with the verbal interactions and they found that males and females tended to adopt different ways of expressing their own ideas. This finding was also confirmed by the findings mentioned in the previous section, suggesting that teachers and researchers need to mix both genders in the future, which may provide a richer learning climate for students to develop their skills of collaboration by working with others with different mindsets. Based on the different definitions of discussion and collaboration (Dillenbourg, 1999), this study found that both gender groups involved different degrees of discussion. Instances of collaboration in boys’ and girls’ groups were also identified. Based on the video observation, the boys individually interacted with the 3D environment and felt free to modify some details of their assigned areas during the whole process. On the other hand, verbal negotiation and reaching a common ground before taking actions seemed to be more emphasized in the female groups. Once the girls had plans fixed, it seemed that fewer adjustments could be made. In this study, male students tended to mix two working styles together. They put their ideas into practice first and then negotiated by directly changing the objects or through further verbal discussion. Female students tended to verbally collaborate in the beginning and then completed their task cooperatively.

It should be noted here that the gender-related differences described in the previous paragraph were based on peer discussion with groups. There were two limitations of this study. One concern was the age of the participants. They were 9 or 10 years old and might have fewer experiences of negotiating in groups. The other limitation relates to the research methods. It was difficult to identify how students negotiated and built their common ground in the groups with a strong leader without exploring the metacognition of their learning experience. Theoretically, more research is needed to explore how young students learn to work collaboratively in different ways, such as well-balanced group distribution (Daradounis et al., 2006). From the practical perspective, teachers can guide students to reflect on the process of their co-creation, especially the process of their collaboration including their verbal communication and actions of their avatars, which could be recorded and reviewed. This metacognition may effectively develop students’ collaboration skills.
The possibility of collaborative learning in a 3D VR world

In this study, the participating students had to construct virtual contexts and create their own stories together. Most of them stated that they learned how to collaborate with others from this experience, for example how to express their own ideas, and to negotiate with someone who held a different idea. These important social skills are not commonly addressed in academic subjects in the education system in Taiwan, which is more teacher-centered. In this study, students were only taught how to create their own contexts in Build & Show and were told to make a short film as the learning outcomes. Although some heated arguments and unfocused behaviors were identified during the process, this study proved that students could learn to collaboratively create their own works even without teachers’ guidance. However, in order to enhance students’ social skills, teachers can set up some events such as an emergency in a 3D VR world and then encourage students to solve the issues while providing certain communication skills. Therefore, students could formulate their concepts of collaboration. With the aid of a 3D VR platform, learning an abstract concept and playing could be done at the same time.

Chang (2009) argued that learning in 3D virtual worlds may increase the workload of learners, in that they need to be familiar with the new techniques and complete learning tasks. However, this was not clearly observed in this study as some researchers (Harasim, 2017; Tapscott & Williams, 2008) identified that the new technology is integrated into the new generations’ life. In this study, more than half of the participating students had experience of playing 3D online games, which means that students could develop their computer skills based on their previous experience. Although the gender differences existed in their creating behaviors and working styles, the female students enjoyed more and were more engaged with their creations in a 3D VR world. This result echoed Robertson’s research (2012). Based on these two studies, it is suggested that teachers should be aware of their unconscious preconceptions of gender bias when deciding whether to incorporate games into their teaching plans. However, according to this study, girls may need more time or more guidance to master operational skills in the beginning. Therefore, the findings suggest that teachers may provide more support for female groups and also guide students on how to deal with the learning difference between members. Although the participating students nowadays are often surrounded by new technology, they still spend a long time trying to control their excitement to focus on their learning tasks. It is suggested that teachers need to carefully develop learning plans with new teaching aids and wisely guide students to the best usage of new technology to enhance their learning, instead of allowing technology to distract them from learning purposes. For researchers, it should be expected that it may take more time to spot participants’ target learning behaviors because it is likely that participants are too excited to stay focused on completing the main tasks. Finally, the sample size was small in this study, so there is a need that a more long-term study dedicating to how a 3D VR world can enhance students’ competency should be done.

Acknowledgments

We would like to thank two anonymous reviewers and editors for their valuable comments and suggestions for improving this article. We thank the Ministry of Science and Technology, Taiwan, R.O.C., under Grant Nos. MOST 105-2511-S-003-018-MY3 and MOST 106-2511-S-003-015-MY3 for financially supporting this research. We are also grateful that this research was partially supported by the “Chinese Language and Technology Center” of National Taiwan Normal University (NTNU) from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education(MOE) in Taiwan.

References


Lee, M. (2009). How can 3D virtual worlds be used to support collaborative learning? An Analysis of cases from the literature. *Journal of E-Learning and Knowledge Society, 5*(1), 149-158.


A Study of the Use of Wearable Devices for Healthy and Enjoyable English as a Foreign Language Learning in Authentic Contexts

Rustam Shadiev¹, Wu-Yuin Hwang² and Tzu-Yu Liu³
¹Nanjing Normal University, Nanjing, Jiangsu Province, China // ²National Central University, Jhongli, Taiwan // ³Bei-Zheng Junior High School, Taipei, Taiwan // rustamsh@gmail.com // wyhwang@cc.ncu.edu.tw // dtyliu@gmail.com

Corresponding author

ABSTRACT
We designed an English as a foreign language (EFL) learning activity supported by smart watches to combine EFL learning with physical exercise such as walking around the school community. We tested the feasibility of our approach to facilitate EFL learning and make it healthy and enjoyable through single subject experimental design. In addition, we researched the affordances of smart watches for EFL learning and healthy and enjoyable living. We collected the following data: (1) student learning outcomes evaluations, (2) a questionnaire survey, and (3) interviews with the students. According to our results, the students performed the best on learning tasks when they used smart watches. The students perceived that smart watches were easy to use and useful for EFL learning. They also perceived that the learning activity supported by smart watches was useful for their health and positive emotions. The students mentioned several features of smart watches that were useful for EFL learning, physical activity, and positive emotions. For example, the dictionary in smart watches helped the students translate unfamiliar vocabulary, and the fitness tracking tool helped track and record the number of steps taken. The students were happy to learn EFL and to monitor their physical activity progress using smart watches. Furthermore, a significant correlation between learning performance and physical activity was revealed, suggesting that the students who did more physical exercise are those who demonstrate better learning performance. On the other hand, no correlation between learning performance and student perceptions and between physical activity and student perceptions was found, suggesting that most students, no matter how well they performed and exercised, had positive perceptions. Based on our results, we suggest designing learning activities combined with physical exercise supported by smart watches to facilitate EFL learning, physical activity, and positive emotions.

Keywords
Authentic contexts, EFL learning, Health, Positive emotions, Wearable device

Introduction

Scholars have suggested that for learning to be meaningful and effective it should take place in authentic contexts (Alioon & Delialioglu, 2017; de la Guia et al., 2016; Herrington & Herrington, 2008; Kiernan & Aizawa, 2004). The reason is because authentic contexts reflect the way the knowledge will be used in real life (Huang, Yang, Chiang, & Su, 2016). Authentic contexts can be found outside of school, e.g., convenience stores or cafeterias in the local community (Huang, Shadiev, Sun, Hwang, & Liu, 2017). Scholars have also suggested that students need to do physical activities to maintain their health and positive emotions (Maclntyre & Vinceze, 2017; Malina, 2001; Pishghadam, Zabetipour, & Aminzade, 2016; Uher, Kuchelova, Cimbolakova, & Pivovarnik, 2016). An instructor may design a learning activity for students to apply newly learned knowledge outdoors and to engage in physical exercise (e.g., walking) at the same time. In turn, physical activity may bring about positive emotions. In this case, the learning activity will facilitate learning, health and positive emotions, and the learning will be healthy and enjoyable.

Due to recent advancements in technological development, it became possible to produce smaller, cheaper and faster computing devices (Shadiev, Hwang, & Huang, 2017). As a result, stationary, portable or mobile computers started to evolve into seamlessly wearable technology featuring an increasing number of smart functions with high-computing power (Kim & Shin, 2015). Wearable technology, such as smart watches, appeals to a broad range of user interests, as it incorporates a wide variety of sensors for continuously measuring, recording and displaying different information, e.g., fitness, health-monitoring, location tracking, voice recognition and recording, and communication (Bower & Sturman, 2015; Müller, Divitini, Mora, Rivera-Pelayo, & Stork, 2015). This is why smart watches can be regarded as potential tools to support learning and healthy and enjoyable living, particularly if these are combined into healthy and enjoyable learning. However, despite the increased interest by the research community in smart watches, few studies have been carried out with such a focus. For example, the effectiveness of smart watches on healthy and enjoyable learning has not been closely examined. In addition, the affordances of smart watches for healthy and enjoyable learning are still unclear. The
results related to the effectiveness of smart watches on and the affordances of smart watches for learning and healthy and enjoyable living can be useful for researchers and educators in the field.

Therefore, this study seeks to obtain data that will help address this research gap. To this end, we designed an EFL learning activity combined with physical exercise and supported by smart watches. During the activity, students applied newly learned knowledge to the real world and physically exercised at the same time. Smart watches helped students complete learning tasks, monitor physical activity, and communicate with other students regarding their healthy and enjoyable learning. We aimed to explore whether our learning activity supported by smart watches can facilitate EFL learning and make it healthy and enjoyable. In addition, we investigated the affordances of smart watches to support healthy and enjoyable learning. Finally, we studied the relationship among the research variables. The following research questions were addressed:

- Do students perform better on a learning task when they use smart watches?
- How do students perceive a learning activity supported by smart watches?
- What are the affordances of smart watches for healthy and enjoyable learning?
- How are the research variables of this study correlated?

Related studies

Learning in authentic environments

Situating learners in real scenarios where they can meaningfully learn and practice has become a major priority for many educators and researchers (de la Guía et al., 2016; Shadiev, Huang, Hwang, & Liu, 2018). When learning is extended to authentic environments, it becomes more effective and meaningful (Alioon & Delialioğlu, 2017; Kiernan & Aizawa, 2004; Shadiev, Huang, Hwang, & Liu, 2017). According to the theory of authentic learning, context and learning should not be separated (Herrington & Herrington, 2008; Shadiev, Hwang, Huang, & Liu, 2015). Students learn much better when they are immersed in real scenarios because their interaction with learning contexts has a profound impact on the way they interpret an activity and their engagement with it (de la Guía et al., 2016). Lave and Wenger (1991) argued that learning is context-related, and they emphasized the importance of learning in a specific context. Collins (1988) suggested that learning is significantly influenced by situations; however, learning in the class is abstract and disconnected from real-life scenarios. That is, schools ignore the interdependence of context, situation and cognition and teach the abstract and decontextualized knowledge, so it cannot be retrieved in real-life contexts (Alioon & Delialioğlu, 2017; Shadiev, Hwang, & Liu, 2018). Under such an approach, knowledge itself is seen by learners as the final product of education rather than a tool to be used dynamically to solve problems.

Herrington and Herrington (2008) listed several critical characteristics of authentic environments. First, authentic environments provide authentic contexts that reflect the way the knowledge will be used in real life. That is, learning should take place in a physical environment containing a large number of resources that preserves the complexity of the real-life setting and reflects the way the knowledge will ultimately be used. Second, authentic environments provide authentic activities. Such activities reflect the kind of activities that people carry out in the real world; they are meaningful and relevant to students and present complex tasks to be completed over a sustained period of time, rather than a series of shorter disconnected examples. Third, authentic environments create opportunities for learners to share their learning experiences and to practice with other learners of various levels of expertise. That is, students share their experiences and are able to access experiences of learners at various levels of expertise. As a result, students learn different perspectives on the topics from various points of view and model their skills and performance based on that of experts. Fourth, authentic environments offer authentic learning assessments within the tasks and promote reflection. The assessment is integrated with the learning activities, peer assessment is encouraged, and learners are assessed based on their outcomes. Learners have the opportunity to compare themselves with other learners at varying stages of accomplishment and improve their performance and skills.

Communicative language teaching approach and task-based learning method

The communicative language teaching (CLT) approach emphasizes interaction. According to this approach, language skills should be developed through meaning-based real communication (Ellis, 2009). That is, learners should learn not only the structure and forms of a language but also its function and purposes and produce language output to advance their communicative abilities (Nunan, 2004). Task-based language learning (TBLL) is one central method of the CLT approach (Kiernan & Aizawa, 2004). This method is student-centered, and it
emphasizes the use of the target language to complete meaningful tasks (Nunan, 2004). In TBLL, the instructor designs different tasks, and the learners use the target language to complete the tasks (Kiernan & Aizawa, 2004). Ellis (2009) argued that such tasks must satisfy the following criteria: (1) language learning activities should focus on “meaning” so that processing the semantic and pragmatic meaning of utterances is emphasized; (2) some kind of knowledge “gap” should exist that requires learners to convey information, to express an opinion or to infer meaning; (3) in order to complete language learning activities, learners should make use of their own linguistic and nonlinguistic resources; and (4) there should be a clearly defined outcome other than the use of language.

Healthy and enjoyable learning

Our lifestyle choices decisively affect the risk of developing major noncommunicable diseases. Physical inactivity is a type of lifestyle behavior with insufficient physical activity when a person spends most of his time sitting or lying, while reading, using a computer, watching television and so on. Physical inactivity is the major risk factor that increases the likelihood of noncommunicable diseases. Such a lifestyle may weaken the immune system, lead to muscle atrophy, cause anxiety and other undesirable health problems (Uher et al., 2016). On the other hand, physical activity enhances or maintains physical fitness and overall health and wellness. Despite the well-known benefits of physical activity, many adults and children are not active enough and lead relatively sedentary lifestyles. Therefore, for the health and well-being of the individuals, the importance of regular physical activity lifestyle throughout childhood, adolescence and adulthood is emphasized, and opportunities to engage in physical activity should be increased (Malina, 2001).

The positive effects of physical exercise on learning performance were reported elsewhere. For example, Salinas, Messias, Morales-Campos, and Parra-Medina (2014) examined the relationship between English language proficiency (ELP), physical activity, and physical activity-related psychosocial measures. Their findings showed that ELP was associated with physical activity. Liu, Sulpizio, Kornpetpanee, and Job (2017) tested for possible positive effects of physical activity when learning a second language (L2). Participants in an experimental group rode a bicycle, whereas participants in a control group sat on a chair during the experiment; all of them were simultaneously presented with new vocabulary and corresponding pictures/translation. The results showed that physical activity improved L2 learning; the experimental participants performed much better on posttests compared to their counterparts. Christopher, Dzakiria, and Mohamed (2012) combined English learning and sports activity. First, in class, students learned about a particular sport (e.g., badminton or basketball), discussed it, and learned related vocabulary. Then, outdoors, the students participated in the sporting activity and used English to interact with the people around them. After the activity, the students described their feelings, their surroundings, the rules and regulations of the sport, their performance, and the skills and actions involved. The results showed that there was improvement in use of the English; students were more expressive in using the English and more confident and motivated after the sports activity. Some scholars also explored the relationship between health and education. Most of them agreed that physical education programs need to be incorporated into curricula to maintain the health and learning performance of students. Two studies showed that students who participated in such programs demonstrated good health and did not experience a drop in standardized test scores (Sallis et al., 1999; Dwyer et al., 1983). Another study showed that students who engaged in physical education programs maintained overall health and wellness and scored significantly higher than their peers who had not engaged in such programs (Schoener et al., 1988). Several studies showed that learning languages has an impact on health. For example, Zauche et al. (2017) suggested that language learning influences education and health outcomes. Tam and Page (2016) claimed that those with good language abilities have better general health. Antoniou and Wright (2017) argued that language learning can be important for maintaining healthy brain function, e.g., working memory, declarative memory, as well as the interaction between declarative and procedural memory.

Emotion is very important in our everyday life, as it is a core process that impacts almost everything we do (MacIntyre & Vincze, 2017). Emotion is the representation of internal states and is tied to physical and sensory feelings (Pishghadam et al., 2016). Emotions can be negative (e.g., anxiety, fear or anger) or positive (e.g., joy and interest). The negative emotional state of students has proved to be disruptive for learning (Fredrickson, 1998). This is especially true during the foreign language learning process because languages are difficult to learn (Pishghadam et al., 2016). Students have negative experiences processing learning material and lower academic success during the learning process, and this arouses intense negative emotions (MacIntyre & Vincze, 2017). According to related studies, negative emotions not only affect learning but also health by causing a number of physical health problems, most notably coronary heart disease and some cancers. On the other hand, the positive emotional state of students has proved to be essential for improving learning. For example, Pekrun,
Goetz, Titz, and Perry (2002) suggested that students’ positive emotions relate in positive ways to student learning, self-regulation, and achievement. Pekrun et al. (2002) found that positive emotions correlated with students’ academic motivation, effort spent at academic tasks, use of metacognitive strategies in learning, and academic achievement. Fredrickson (1998) and Pishghadam et al. (2016) argued that positive emotions broaden the scope of attention, cognition, and action as well as build physical, intellectual, and social resources. That is, those in a positive emotional state notice more items in their visual field, engage more social connections, and tend to have urges to act in a greater variety of ways relative to those with negative emotions (MacIntyre & Vincze, 2017). According to Tugade, Fredrickson, and Barrett (2004), positive emotions are good for health. Furthermore, positive emotions contribute to psychological and physical well-being (Tugade, Fredrickson, & Barrett, 2004). For example, positive emotions are useful in preventing and treating anxiety and depression and thereby optimizing health (Fredrickson, 1998). Other scholars have argued about the impact of exposure to physical activity on emotions. For example, Biddle (2000) claimed that performing physical activity has positive benefits on emotions and moods. Wankel and Berger (1990) found the psychological and social benefits of physical activity; positive emotions were reported to be a main reason for physical activity involvement. Therefore, it is important to cultivate positive emotions in students by engaging them in different physical activities.

To make students learn better, be physically active and happy, a learning activity can be designed. For example, teachers may incorporate learning and physical exercise (e.g., walking to and from school instead of taking a bus) into a learning activity. Such an opportunity can be created to weave physical exercise into students’ daily routines with respect to many different subjects. For a language learning class, following task-based language learning and authentic learning principles, students can be assigned learning tasks to apply newly learned knowledge to the real world. Teachers may ask students to explain to their classmates how to get to their home from school in the target language. Students may complete this task on their way home from school, and objects, situations and scenarios in the surrounding contexts can help student complete their tasks. Walking while working on the learning task may bring about positive emotions. Such a learning process will lead to healthy and enjoyable learning.

### Smart watches for healthy living and learning

A smart watch is a wrist-worn, general-purpose, networked computer with a wide variety of sensors for continuously measuring and displaying different information for users (Barfield & Caudell, 2001; Reeder & David, 2016). Smart watches are familiar to most people, increasingly available as a consumer device, and enable near-real time continuous monitoring of physical activity and physiological measures. Mukhopadhyay (2015) claimed that smart watches are very popular in many applications, such as medical, entertainment, education, and other fields. According to recent studies (Lee, Bojanova, & Suder, 2015), health and activity monitoring were the most popular uses for wearable technologies, followed by information access (i.e., tell time, search for directions) and communications (i.e., send a text). The following example may explain how smart watches can be useful to support healthy living. With respect to daily physical activities, smart watches display a feedback with the number of steps taken, and users may compare it to the standards of their personal goal. Such feedback has positive effects on goal engagement. On one hand, if a user’s number of steps has contributed to accomplishing the goal, the feedback shows that user is on the right track and motivates continued action. On the other hand, if the goal was not accomplished, feedback signals that more effort or self-control is needed. Therefore, feedback motivates users to take enough or more steps, and it stimulates and supports healthy behavior (Dijkstra & Kooy, 2017).

Smart watches can also be successfully applied to education (Dijkstra & Kooy, 2017; Mukhopadhyay, 2015). The reason is because smart watches are useful and provide a wide range of opportunities to educators due to the following pedagogical affordances (Bower & Sturman, 2015). Smart watch devices have the ability to provide in situ contextual information. That is, smart watches can be used for receiving or searching for additional background information. The recording of information using smart watches was also identified as a pedagogically benefit, e.g., students can use smart watches to record and show what they are doing in the learning contexts as well as to replay recorded content for reflective reprocessing. Communication features include the opportunity to integrate communication streams into the learning process to make it social, which is very important because students across different locations may exchange learning information, discuss ideas with others, help one another to complete class assignments, or see an event from someone else’s viewpoint (Wu et al., 2014). As a result, this can be useful in increasing the acquisition and retention of new knowledge. Smart watches are seen to provide timely unobtrusive and contextualized feedback, e.g., students can receive instant feedback on their recorded content from the system or from other students via online chat and then try to
improve their content. In addition, smart watches provide well-timed and relevant push notifications to reach students when they are busy in order to grab their attention and facilitate healthy living and/or learning, e.g., to suggest they take more steps or to remind them of a newly learned concept. The hands-free access of smart watch devices enhances learning experiences through logistical and other pragmatic implications, e.g., smart watches provide learners with hands-free access to contextually relevant knowledge (Bower & Sturman, 2015), and learners are able to send or receive in situ information and keep moving at the same time.

Lungu (2016) argued that the smart watch is a useful learning tool in a way that makes a user consult it easily for a learning strategy called microlearning, which is known for quickly closing skill and knowledge gaps. The learner can take advantage of the smart watch and study anytime and anywhere, especially during dead moments of the day (e.g., waiting for the school bus at the bust stop). Chen (2016) suggested that learning in smart watch-enhanced learning environments takes place in a state of physical mobility. Such environments are different from those created by mobile technologies (i.e., smart phones) because the devices to be used in former are wrist-worn, which enables learners to interact with them without slowing down their movement or standing still. That is, wearable technology enables learners to move and interact with the device smoothly.

Recent evidence suggests that there have been few studies that have been sufficiently conducted to explore smart watch applications for learning, or the relevant studies are still in the preliminary stages (Bower & Sturman, 2015; Kim & Shin, 2015). However, some learning scenarios were proposed for smart watch-supported learning environments. Dijkstra and Kooy (2017) suggested that learners may receive questions or assignments and feedback on their learning via smart watches. For example, secondary school students may be presented with some familiar routes in their environment as well as unfamiliar ones and be asked to calculate how many steps they need to take for each. The accelerometer sensor of smart watches detects users’ steps and counts them (Reeder & David, 2016); this sensor may provide feedback containing social comparisons and competition regarding physical activity, e.g., “you are doing great compared to your friends!” Buchem, Merceron, Kreutel, Haessner, and Steinert (2015) proposed extending a typical learning environment with the level of engagement made available through smart watches. Smart watches can capture live data from individual learning activities and identify appropriate moments for carrying out learning interventions in a wide variety of scenarios. Chen (2016) used smart watches for simple, unobtrusive physiological measurements. Smart watches measured and tracked heart rate changes during self-paced learning processes and provided students with dynamic feedback about their emotional status in order to alleviate their anxiety and promote learning engagement.

### Language learning with smart watches

Smart watches may offer various pedagogical benefits for language learning. First, smart watches incorporate wireless connectivity and provide seamless access to contextually relevant information as well as enable interaction with information (Bower & Sturman, 2015). This makes smart watches a valuable tool to aid language learning both in the classroom and outside of it (Kim & Shin, 2015). Second, the multimedia tools of smart watches allow students to create their own content in environments outside of school and access content created by other students (Sen, Subbaraju, Misra, Lee, & Balan, 2016). For example, students can practice their writing or speaking skills by introducing or describing some objects, situations and scenarios from surrounding contexts by creating texts, images, and audio files using multimedia tools of smart watches. Students then can share their content with peers using smart watches for further reflection, discussion, and collaboration (Dubey, Goldberg, Abtahi, Mahler, & Mankodiya, 2015). Another feature of smart watches is feedback or notifications that are generated as a visual cue, auditory signal or haptic alert. Students can receive optimally timed notifications of newly learned concepts, and their repetitions can lead to language skills improvement (Lungu, 2016).

Some studies on EFL learning using smart watches were carried out recently. An English practice application has been developed with a card-based interface for smart watches in Pham, Chen, Nguyen, and Hwang (2016). Since only a few users used this application on smart watches, the researchers focused on its effectiveness for personalized adaptive learning with smartphones. Lungu (2016) developed the Time to Learn application for the smart watch to learn foreign vocabulary. The application was readily available anytime a student looked at it. It presented a word recognition challenge for the student, and after every challenge, the student provided feedback on whether he/she knew a given word or not. These two studies focused on vocabulary learning and were still preliminary. de la Guia et al. (2016) introduced IoT and wearable technologies for young students learning a foreign language. Students were asked to prepare a meal; they shared the responsibility for finding the various ingredients, shopping, and finding the appliances and kitchenware to lay the table. The students were provided with many IoT objects (e.g., food products and kitchenware items) with near-field communication (NFC) tags,
and they had to identify a given item from among a large number of objects and scan it via the NFC reader on the wearable device. The results of their study showed the great benefits of using wearable and IoT technologies for creating realistic scenarios in which learners can meaningfully and effectively learn a foreign language.

**Method**

For this study, we employed an experiment. According to Fraenkel, Wallen, and Hyun (2014), intervention can be tested with an experiment to determine whether it influences outcomes. We employed single subject experimental design. According to Fraenkel et al. (2014), such a design is most commonly used to study the changes in behavior and performance an individual exhibits after exposure to an intervention. We followed an A-B principal method of single-subject experimental design; this is a two phase method composed of an intervention condition (Phase A) and a control condition (Phase B) (Creswell, 2014). If any changes are observed, then one may conclude that the intervention has had an effect. The usefulness of such a research design and its reliability and validity are reported in Creswell (2014) and Fraenkel et al. (2014). In this present study, students participated in a learning activity supported by smart watches (Phase A: an intervention condition) first and then they participated at a learning activity without smart watches support (Phase B: a control condition). We compared students’ learning outcomes obtained after each phase to prove that a learning activity supported by smart watches is feasible and effective.

**Subjects and research procedure**

Eighteen junior high school students from one class participated in this study. The demographic information of the participants is presented in Table 1. Most of the participants were 14-15 years old. Half of the participants were boys, and the other half were girls. This is because equal numbers of boys and girls are usually assigned in classes of most Asian countries, e.g., China (The State Council Information Office of the People’s Republic of China, 2015). The participants had at least four years of experience using computers, and most of them had used smart watches for less than one year.

<table>
<thead>
<tr>
<th>Table 1. Participants’ profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category</td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Age (years)</td>
</tr>
<tr>
<td>14</td>
</tr>
<tr>
<td>15</td>
</tr>
<tr>
<td>Experience to use computer (years)</td>
</tr>
<tr>
<td>1-3</td>
</tr>
<tr>
<td>4-6</td>
</tr>
<tr>
<td>7 and more</td>
</tr>
<tr>
<td>Experience to use smart watch (years)</td>
</tr>
<tr>
<td>less than 1</td>
</tr>
<tr>
<td>1 to 3</td>
</tr>
<tr>
<td>more than 3</td>
</tr>
</tbody>
</table>

Figure 1 presents our research procedure. We carried out a pretest before the learning activity to measure student prior knowledge. After that, each student received a smart watch, was instructed in how to use it, and was told about the learning activity details. To get acquainted with the smart watch and its functions, the students used it for one week. In the following two weeks, the students participated in a learning activity. The learning activity was supported by smart watches during the first week, and no smart watch support was provided during the learning activity in the second week. In the last class, interviews and a questionnaire were administered to investigate student learning experiences and perceptions.

We admit that this was a short-term study. Fraenkel et al. (2014) argued that, in some studies, the data on variables are usually collected within a fairly short time. In such cases, the instruments used are administered in a single session or in two sessions one immediately after the other. Creswell (2014) suggested such a design for researchers to better control conditions in order to focus on the effect of an intervening variable. This is often easier to do when conducting a short-term investigation. Large scale studies are then carried out to verify the
results from short-term studies. The notion of a short-term intervention has been widely used in educational research, including for mobile-assisted language learning. Here, two examples are provided: (1) the effectiveness of English practice application on EFL learning was explored for a seven-day period in Pham et al. (2016), and (2) the intervention to measure the effects of a situated mobile learning approach on the learning motivation and performance of EFL students was 5 hours long in Huang et al. (2016).

Learning activity

An English course was conducted in this study. A one-hour class was carried out three times a week in a conventional classroom by the instructor. Our learning activity was divided into indoor (in classroom) and outdoor (outside of school) activities. The learning goal of the indoor learning was to improve student vocabulary and reading and listening skills using textbook material, whereas the outdoor learning was aimed at improving student vocabulary and writing and speaking skills by applying newly learned knowledge to the real world. Indoor learning included learning new vocabulary and grammar, reading textbook dialogues, listening to audio dialogues, and then applying newly learned knowledge to complete textbook exercises. Outdoor learning included applying newly learned knowledge to the real world, and for that, the students kept a diary in which they recorded the everyday events that happened to them. The students were asked to think about what happened each day and then to write it down and talk about it. For example, they described objects or people (e.g., a kind and professional waiter in a breakfast store), situations (e.g., encountering a friend in local supermarket), and scenarios/events (e.g., celebrating grandma’s birthday party) they saw or experienced in their local community. More details about the pedagogical benefits of a diary in the learning process can be found elsewhere (Hwang, Hsu, Shadiev, Chang, & Huang, 2015). This learning task was designed based on the “I was sleeping when you called” topic from the course textbook that students learned in class. From the topic, students learned new vocabulary, past continuous and past simple tenses, and sentence patterns of the theme. Then, to complete the task, the students used the newly acquired knowledge to describe what happened to them each day. In the first week, the students wrote entries in a diary with the support of smart watches (Task 1), and in the second week, the students kept a diary without smart watch support (Task 2). The screen size of smart watches is too small and, therefore, may be too inconvenient when inputting diary entries, so we asked the students to keep paper-based diaries and write entries in those diaries. Smart watches were used by the students to aid them in keeping a diary. For example, they used the (1) dictionary to translate new vocabulary for a diary and to determine how this vocabulary can be used in different contexts; (2) the communication tools to communicate with their peers, for example, to call their classmates to request/provide help or discuss some other learning issues; (3) the voice recording tool to record verbal content (e.g., description of objects, people, situations and scenarios) and share with peers for further discussion; and (4) the speech recognition tool to speak in English and get feedback about pronunciation. In addition, the students used (5) the fitness tracking tool of the smart watches to monitor their physical activity.

Research tools

We collected and analyzed the following data: (1) pretest and healthy learning tasks, (2) perceptions toward the learning activity supported by the smart watches, and (3) interviews with students. The use of multiple data sources allowed for the triangulation of the main findings and rendered the conclusions richer, more nuanced, and more reliable.
Pretest and healthy learning tasks

In the pretest, the students were asked to recall what has happened in their life lately and write it down. For example, the students could write about an important event, such as having an important conversation with their friends/parents/teachers or seeing something unusual on their way back home from school, and their feelings about it. We designed two healthy learning tasks for students (please refer to the Learning activity section). For the test and tasks, the students were asked to use appropriate sentences (at least 3-5 sentences) and the sentence patterns they learned in class. In terms of the healthy aspect of the learning tasks, we asked the students to be physically active (i.e., walk around their local community) when they worked on the tasks. The smart watch displayed how many steps they took every day, and the students included this information in their task sheets.

Perceptions toward the learning activity supported by smart watches

We employed a questionnaire survey to explore student perceptions toward the learning activity supported by smart watches. The questionnaire was developed based on the technology acceptance model (TAM) proposed by Davis (1989). The TAM has been successfully employed in a wide array of educational research areas (Hwang et al., 2014; Hwang et al., 2016). Researchers demonstrated that a TAM-based questionnaire is valid and reliably measures learners’ perceptions toward the intervention (Davis, 1989). The following dimensions were included in the questionnaire: ease of smart watch use or EU (four items), the degree to which a student believes that using a smart watch would be free of physical and mental effort; the usefulness of the learning activity supported by the smart watch or UL (four items), the degree to which a student believes that the learning activity supported by the smart watch would enhance his or her learning performance; behavioral intention to use a smart watch for learning in the future or BI (two items), a learner’s subjective probability that he or she will use a smart watch for learning in the future; the usefulness of the learning activity supported by smart watches for health or UH (three items), the degree to which a student believes that the learning activity supported by the smart watch would enhance his or her health; and the usefulness of the learning activity supported by smart watches for positive emotions or UPE (two items), the degree to which a student believes that the learning activity supported by the smart watch would enhance his or her positive emotions.

Interviews

We conducted one-on-one semistructured interviews with the students. The interviews were aimed toward an exploration of the students’ experiences using the smart watches during Task 1 and toward the provision of insights into their perceptions about the usefulness of the learning activity supported by smart watches for learning, health, and positive emotions. Each interview lasted 20 minutes.

Smart watch

We employed ASUS ZenWatch 2 smart watches (Figure 2-a) to support healthy and enjoyable student learning (ASUS, 2017). The ASUS ZenWatch 2 features a 1.63 inch 320x320 touch display, a Qualcomm Snapdragon 400 processor, and 512 MB RAM and 4GB ROM. It comes with Bluetooth and WiFi connectivity features to connect to the internet or other devices. The ASUS ZenWatch 2 is compatible with many applications, e.g., Wellness App. The ASUS ZenWatch 2 has a built-in speaker and built-in microphone that enable making and receiving calls (Figure 2-b), setting audible alarms, and recording and hearing app notifications and sounds (Figure 2-c). In addition, users can exchange short messages, emojis, and multimedia files with others (Figure 2-d) and view different notifications and other important information on a display. A built-in pedometer counts...
steps and reports activity progress in real time as well as daily and weekly activity summaries in the form of timelines and charts (Figure 2-e). In addition, we developed the ezTranslate APP for smart watches (Figure 2-f), which enables speech input through speech-to-text recognition technology, the translation of vocabulary from English into Chinese and vice versa, and saying vocabulary or sentences aloud through text-to-speech recognition technology.

Data analysis

The written content of the task outcomes was coded using a sentence as a coding unit. We scored them on a 10-point scale (with 10 as the highest score). The following assessment criteria were applied (The Ministry of Education, 2007): (1) Knowledge and Understanding: subject-specific content acquired, and the comprehension of its meaning and significance; (2) Thinking: the use of critical and creative thinking skills and/or processes; (3) Communication: the conveying of meaning through various forms; and (4) Application: the use of knowledge and skills to make connections within and between various contexts. Task outcomes were assessed in a balanced manner with respect to the four categories. Student speaking performance was also measured with respect to pronunciation and fluency. To this end, the students had to speak their created content aloud. Differences in learning performance were measured by comparing student learning performance on each task. The assessment was carried out by three raters, and notable differences in the scoring were resolved through rater discussions until a consensus was achieved. The inter-rater reliability of the scoring was evaluated using Cohen’s kappa. The result exceeded 0.80 (before the discussion) and 0.90 (after the discussion).

Eighteen valid answer sheets to the questionnaire were obtained from 18 students. The students responded to the questionnaire items using a five-point Likert scale anchored by the end-points “strongly disagree” (1) and “strongly agree” (5). The internal consistency of the survey was tested by employing Cronbach’s α. The values for EU (α = 0.827), UL (α = 0.941), BI (α = 0.833), UH (α = 0.975), and UPE (α = 0.873) demonstrated the high reliability of the questionnaire.

All interviews were audio-recorded with the students’ permission and were then fully transcribed for analysis. The text segments that met the criteria to provide the best research information were highlighted and coded. The codes were then sorted into categories; codes with similar meanings were aggregated together. Established categories formed a framework to report findings pertinent to the research questions. The inter-rater reliability of the interview data was also evaluated by using Cohen’s kappa, and the result exceeded 0.90.

A paired samples t-test was employed to measure the differences in the student learning performance on the pretest and the tasks completed with and without the smart watches. Paired samples t-tests typically consist of one group of units that has been tested two or more times. We also computed the correlation among our research variables to measure the degree to which they are related. To this end, we employed the Pearson correlation coefficient. The a priori alpha-level was set at 0.05 for all statistical analyses in this study since an alpha level of less than 0.05 is acceptable in most educational research as statistically significant.

Results

Prior knowledge and learning performance with and without smart watches

The results of the student prior knowledge and learning performance evaluation on the tasks with and without smart watches are presented in Table 2. According to the results, there was a significant difference between scores on the pretest (M = 2.89; SD = 2.72) and learning performance on Task 1 (M = 6.17; SD = 2.06), t = -5.804, p = .000. In addition, the results showed a significant difference between scores on the pretest (M=2.89; SD=2.72) and learning performance on Task 2 (M = 5.61; SD = 2.77), t = -4.151, p = .001. Finally, we found that learning performance on Task 1 (with smart watches) was significantly higher than learning performance on Task 2 (without smart watches), t = 2.149, p = .046.

Table 2. Results of the pre-test and learning performance evaluation

<table>
<thead>
<tr>
<th>Prior knowledge</th>
<th>Task 1 (with smart watches)</th>
<th>Task 2 (without smart watches)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>2.89</td>
<td>2.72</td>
<td>6.17</td>
</tr>
<tr>
<td>5.61</td>
<td>2.77</td>
<td></td>
</tr>
</tbody>
</table>
Student perceptions toward the learning activity supported by smart watches

The results of the questionnaire survey are presented in Table 3. According to the results, student perceptions were high regarding the ease of smart watch use ($M = 4.26$, $SD = 0.56$), the usefulness of the learning activity supported by smart watches for learning ($M = 3.99$, $SD = 0.74$), the behavioral intentions to use smart watches for learning in the future ($M = 4.47$, $SD = 0.51$), the usefulness of the learning activity supported by smart watches for health ($M = 4.57$, $SD = 0.50$), and the usefulness of the learning activity supported by smart watches for positive emotions ($M = 4.67$, $SD = 0.48$).

Table 3. Student perceptions toward the learning activity supported by smart watches

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ease of smart watches use</td>
<td>4.26</td>
<td>0.56</td>
</tr>
<tr>
<td>Usefulness of the learning activity supported by smart watches for learning</td>
<td>3.99</td>
<td>0.74</td>
</tr>
<tr>
<td>Behavioral intentions to use smart watches for learning in the future</td>
<td>4.47</td>
<td>0.51</td>
</tr>
<tr>
<td>Usefulness of the learning activity supported by smart watches for health</td>
<td>4.57</td>
<td>0.50</td>
</tr>
<tr>
<td>Usefulness of the learning activity supported by smart watches for positive emotions</td>
<td>4.67</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Affordances of smart watches for healthy and enjoyable learning

Based on the interview results, the following affordances of smart watches for healthy and enjoyable learning were derived: hands free access, translation, speech- and text-to-speech recognition, notifications, voice recording, information sharing, communication, fitness tracking, and enjoyable learning.

Correlation among research variables

We explored the relationship among the following research variables: (1) scores for Task 1, (2) steps taken, and (3) student perceptions. According to our results, the scores for Task 1 significantly correlate with the steps taken ($r = 0.652$, $p = .003$). However, no significant relationship was found between the scores for Task 1 and student perceptions as well as between steps taken and student perceptions.

Discussion

The effectiveness of the learning activity supported by smart watches

Our results suggest that the students gained new knowledge during the learning activities, so their performance on the tasks with and without smart watches was higher compared to their prior knowledge. The difference between student learning performance on Task 1 (with smart watches) and Task 2 (without smart watches) suggests that smart watches were useful for significantly promoting student EFL learning. This finding is in line with those obtained in earlier related studies. For example, several scholars claimed that using wearable devices, such as smart watches, during language learning can facilitate language practice in authentic contexts and, therefore, foster learning outcomes (de la Guía et al., 2016; Lungu, 2016; Pham et al., 2016). The reason is because smart watches offer several pedagogical benefits, e.g., wireless connectivity so that students have seamless access to learning information, and they can interact with their teacher and peers (Bower & Sturman, 2015; Kim & Shin, 2015). In addition, students are able to create their own content using the multimedia tools of smart watches (Dubey et al., 2015; Sen et al., 2016). Furthermore, students can receive feedback/notifications on the learning material they need to study (Lungu, 2016). These features of smart watches enable the students to practice the language frequently.

Student perceptions

Our results suggest that, in general, the students had positive perceptions toward the learning activity supported by smart watches. This finding was supported by the interview results; during the interviews, the students again confirmed their positive perceptions toward the learning activity supported by smart watches. All the students agreed that the learning activity was useful for EFL learning, they took more steps during the learning activity, and it was useful for their health and positive emotions. Smart watches helped them participate in the learning activity more efficiently.
Related studies have suggested that the acceptance of technology should be evaluated on a pedagogical basis to interpret its usage (Hwang et al., 2014; Hwang et al., 2016). Davis (1989) suggested that technology acceptance can be measured through perceived ease of use, perceived usefulness and behavioral intention dimensions. Our results suggest that the students accepted smart watches for language learning in terms of its ease of use and usefulness as well as their behavioral intentions. Furthermore, the students accepted that our learning activity is useful for learning, health and positive emotions.

**Affordances of smart watches**

Our results revealed the following affordances of smart watches for healthy and enjoyable learning.

*Hands free access:* The students mentioned that they were able to operate the applications of the smart watch (e.g., open/close applications and scroll down/up applications) with hand gestures. For example, by flicking the wrist in and out, the students could scroll up and down displayed information, or by pushing their arms down, they could select and open applications. This was very convenient during the learning process and physical exercise, especially when students could not use both hands to control the smart watch.

*Translation:* The students said that the smart watch dictionary helped them translate unfamiliar vocabulary from Chinese into English and vice versa. This feature was especially useful when they were working on their task and did not know some vocabulary. In addition, the dictionary helped the students to see what vocabulary they translated recently as well as pin frequently used/important vocabulary in the dictionary.

*Speech- and text-to-speech recognition:* The speech recognition tool was useful for practicing EFL speaking skills. If the students made mistakes during speaking, the speech recognition tool could generate text with errors so that the students knew that they had made mistakes. When the speech recognition tool generated text correctly the students knew that their speech was correct. In addition, the text-to-speech recognition tool spoke out target vocabulary words and sentences so that the students could hear how to say them correctly. The students said that this tool was useful for improving their EFL pronunciation.

*Notifying:* The notifications feature of the smart watch was employed for learning vocabulary. New vocabulary was shown to the students regularly, and this helped them better remember it for a longer time. In terms of physical exercise, the notifications enabled the students to see data related to their steps taken, calories burned, distance traveled, and so on, which were tracked and recorded by the fitness tracking tool. If a student did not reach his/her daily goal, smart watch alerts could pop up and encourage him/her to reach it.

*Voice recording:* The students said that with the smart watch they were able to record their speech. Later, a student could listen to his/her recorded speech again to check whether he/she made any mistakes.

*Information sharing:* The students were able to share their created content, such as a recorded audio file, with other students. This feature was helpful for peer learning. Students with better proficiency could find mistakes in the audio files of lower proficiency students, point mistakes out, and suggest how to improve. In addition, lower proficiency students could listen to the audio files of higher proficiency students and learn from them, e.g., to get inspirational ideas on how to complete the task. With regard to physical exercise, the students shared the goals they achieved with their peers. Seeing the results from others motivated the students to engage in more physical activities. The students tried to take more steps if they saw that their own progress was lower compared to that of others.

*Communication:* The students mentioned that the communication tools of the smart watches, e.g., text or voice messaging, enabled them to communicate with each other. The communication tools were useful for learning by providing assistance to those in need to complete the task or solve related issues.

*Fitness tracking:* The fitness tracking tool tracked and recorded data related to the students’ steps taken, calories burned, distance traveled, workouts completed (e.g., walk, run, cycling, push-ups, sit-ups), sleeping habits and heart rate.

*Enjoyable learning:* The students agreed that smart watches were useful for making them happy. The smart watch functions enabled the students to practice EFL skills and communicate in EFL with less anxiety about making mistakes. Since the students did not communicate with others face-to-face but via the smart watch, they had no pressure of being judged if they made mistakes. Furthermore, according to the students, using smart
watches made the learning process more interesting, fun, and engaging. Regarding physical exercise, the students felt satisfied and happy when they saw notifications about them reaching or surpassing their daily goals. During communication with others, the students cheered each other up and so this made them happier.

**Correlation among research variables**

A significant correlation between the number of steps taken and the scores for Task 1 implies that the students with better performance are those who took more steps. That is, the more steps students took, better they performed on the task. This is because these students went out to complete the task and to do physical exercise and spent more time on both learning and physical activity. They learned and practiced their EFL skills longer, and this is why they performed better. Students, who did not go out so often or for a long time did not practice their EFL skills, and so, their performance was lower. Therefore, we suggest that students need to be encouraged to go out frequently and for longer periods of time to learn and exercise. The reason for the insignificant correlation between the scores for Task 1 and student perceptions and between the number of steps taken and student perceptions is that most students, regardless of whether their scores for Task 1 were high or low and regardless of how many steps they took, believed that it was easy to use the smart watch, that the learning activity supported by the smart watch was useful for learning, that the student had high behavioral intention to use the smart watch in the future, and that the learning activity supported by the smart watch was useful for health and positive emotions.

**Conclusion**

We collected multiple data sources and triangulated the findings to make the research more rigorous. Based on our findings, the answer to the first research question is that the students completed the learning task better when they used smart watches. The answer to the second research question is that the students positively perceived the learning activity supported by the smart watch as facilitating healthy and enjoyable learning. The answer to the third research question is that there are several affordances of smart watches for healthy learning: hands free access, translation, speech- and text-to-speech recognition, notifications, voice recording, information sharing, communication, fitness tracking, and enjoyable learning. The answer to the last research question is that student learning performance correlates with the number of steps taken; however, student perceptions correlate neither with learning performance nor with the number of steps taken.

We make two suggestions for the teaching and research community in the field. First, we suggest designing learning activities supported by smart watches in which students are able to learn new concepts and apply newly learned knowledge to the real world while physically exercising. In this way, enjoyable and healthy EFL learning will be facilitated. Second, instructors need to make sure that students are aware of what affordances smart watches have for enjoyable and healthy EFL learning. This will help the students to utilize smart watches for more efficient, enjoyable and healthy EFL learning.

Several limitations of this study need to be acknowledged and addressed in the future. The first limitation relates to the small sample size involved in this present study. The second limitation is that the students were exposed to the learning activity supported by smart watches for a short time period. As a result, these issues may limit the generalization of the obtained results to the wider population. Another limitation is that, although we used an experimental design, we compared the effects of two different conditions (with and without smart watches) on the learning outcomes of students in the same group. That is, a control group (i.e., with different students) was absent in this study. Thus, it is possible that some other variables could influence our results. For example, the learning motivation of the students could increase during Task 1 after they receive smart watches and result in better learning performance. On the other hand, after the smart watches were taken away, the students could have lower learning motivation during Task 2 and, as a result, perform worse. Therefore, all these limitations need to be considered and addressed in future studies. We will also try to compare the smart watches used in this present study with other smart watches/wearable devices in the near future to identify the differences in terms of their features to support authentic learning with respect to pedagogical, physiological and psychological aspects.

**References**


Investigating Flipped Classroom and Problem-based Learning in a Programming Module for Computing Conversion Course

Adriana E. Chis, Arghir-Nicolae Moldovan, Lisa Murphy, Pramod Pathak and Cristina Hava Muntean

School of Computing, National College of Ireland, Dublin, Ireland // adriana.chis@ncirl.ie // arghir.moldovan@ncirl.ie // lisa.murphy@ncirl.ie // pramod.pathak@ncirl.ie // cristina.muntean@ncirl.ie

*Corresponding author

ABSTRACT
This research paper investigates the effectiveness of combined Flipped Classroom (FC) and Problem-based Learning (PBL) teaching approach in a computer programming module. Combined FC-PBL makes use of learning technologies and supports authentic learning in terms of authentic context, multiple perspectives through team work and collaboration. FC involves watching educational videos prior to class session, and using the class session for practical tasks, thus supporting programming skills development. PBL enables authentic learning activities through group work and helps students to develop 21st century skills such as self-direction, collaboration, creativity and innovation. A research case study was conducted that considered a three-stage based delivery of the module that involved traditional, only FC, and combined FC-PBL teaching approaches applied on a number of programming topics. Both the educational and edutainment benefits were analyzed. The results show that combined FC-PBL approach is effective and the knowledge acquired by students improves, in particular, for the weaker students. Only 1.9% of students have scored an assessment mark lower than 40% compared to 28.3% and 24.5% when the traditional teaching and FC-only approaches respectively were used. Overall, a 26.56% increase in the assessment results was noticed when combined FC-PBL was used. The edutainment element was investigated through a questionnaire that assessed learning environment, engagement, learner satisfaction and whether the students have enjoyed learning sessions. The survey analysis shows that the combined FC-PBL approach does aid in the edutainment of mature students and provides an enjoyable learning experience.

Keywords
Flipped classroom, context-based learning technology, problem-based learning, authentic learning, fun e-learning, edutainment, authentic context

Introduction
Edutainment is marriage of education and entertainment and specifically refers to entertainment that is designed to be educational (Aksalal, 2015). An edutainment-based pedagogy needs to be both educational and engaging/enjoyable. Edutainment is very relevant for education today as it aims to provide education with engagement. An important element is that the entertainment should not reduce the value of education and still meet the learning objectives.

Edutainment is especially relevant for subjects such as computer science and mathematics. In computer science education innovative pedagogies are really needed to engage students in subjects like programming as computer science topics (including programming) are not taught at secondary/high school level in many European countries and therefore, students do not have any background knowledge (Vlatko, 2015; Kebritchi, Ah Hamid & Fung, 2007; Hirumi & Bai, 2010; Passey, 2017). As programming is a new area for students in the college, they need to spend a lot of time solving programming problems and the classroom sessions are not able to fully support this. To learn programming concepts and to develop good programming skills one needs to learn to extract the algorithm from the requirements and then translate the algorithm to software code using the syntax of a particular language. To accomplish this a student needs to start with basic problems and programming tasks and gradually practice to complex problems (Winslow, 1996). Various approaches and pedagogical methods have been explored by researchers and educators to overcome the challenges with teaching and learning programming modules. A good approach to deal with programming is to motivate the students by using edutainment-based pedagogical strategies that may involve problem solving practical approaches, authentic context showing how the acquired knowledge will be used in real life, conceptual learning, collaborative learning, authentic activities, self-directed and active learning, practice and learning from failure, increased practical hours.

A relevant pedagogy for edutainment and engagement is constructivism (including socio-constructivism). Constructivist theory of education says that people construct their own understanding and knowledge through experience and reflection. The constructivist pedagogy is based on providing students the opportunities to use active learning techniques to create knowledge (real-world problem solving) and to construct their knowledge...
through discussion, collaboration and reflection (socio-constructivism). The constructivist pedagogy is suitable for teaching programming and can be quite effective (Wang, Dong & Li, 2012). Constructivism theory also believes in experiential learning which is based on the design of practical labs (kinaesthetic activities). Technology based pedagogies such as game based learning (GbL), flipped classroom (FC) and problem-based learning (PBL) are constructivist approaches that are based on providing education with edutainment.

GbL involves the use of gaming technology for educative purposes where students explore concepts in a learning context designed by teachers. Considering the advantages of games which are fun and engaging, combined with educational field, educational games make students to participate actively in an edutainment environment and help them learn while having fun. Previous research works have shown that game-based learning can have positive effects on important educational factors such as student motivation and engagement (Ghergulescu & Muntean, 2012; Muntean, El Mawas, Bradford & Pathak, 2018), learning effectiveness (Erhel & Jamet, 2013; Zhao, Chis, Muntean, & Muntean, 2018), as well as learning attitude, achievement and self-efficacy (Sung & Hwang, 2013). Moreover, game-based learning has the potential to facilitate the acquisition of 21st century skills such as critical thinking, collaboration, creativity and communication (Qian & Clark, 2016).

FC is a student-centered pedagogy in which students’ complete pre-class work (e.g., watching a video clip, searching online for certain information) to gain basic knowledge, and class time is dedicated to activities that promote application and mastery of this knowledge. FC is based on the principles of edutainment as it makes classroom an active learning environment. Lectures are conducted in video format outside of the classroom time and this time is then devoted to discussions, collective and individual feedback, reflections, collaborations, problem-solving, etc. The recorded lecture/video can be completed by students at their own pace. These elements make the flipped classroom approach fun and engaging. The concept of flipped classroom, in association with edutainment, was studied by Guy and Marquis (2016), when students were assigned video lessons and podcasts as opposed to projects prior to class. It was found that these students outperformed those in traditional classrooms, found the actual time in class to prompt more interaction, and thought the class to be more enjoyable, although there was a noticeable period of adaption (Guy & Marquis, 2016). FC also supports the development of new skills such as critical thinking, creativity, communication and collaboration.

PBL is a constructivist pedagogy based on hands-on, active learning collaborative approach that helps students achieve their learning objectives by working on a real-world problem. Due to the active nature of this pedagogy, it focuses on engagement and collaboration while solving problems. PBL empowers learners to conduct research and learn by solving real-world and/or open-ended problems. PBL helps students to develop 21st century skills such as self-direction, collaboration, creativity and innovation, as it puts emphasis on student independence and enquiry, and focuses on students working in groups to find solutions to authentic real-life problems (El Mawas & Muntean, 2018; Larmer, 2014; Rotherham & Willingham, 2010).

Interpreting new information in the context of where and when it occurs, and being able to relate it to what we already know, also develop a better understanding of its relevance and meaning. Therefore, the potential of innovative technologies applied in edutainment such as augmented reality (AR), virtual reality (VR) and virtual labs (VL) has been explored in recent years. These innovative teaching approaches also provide an engaging and enjoyable learning experience and support learning through practicing, hands-on activities and reflection.

AR can engage, educate and entertain learners in a new way. AR adds information or meaning to a real object, integrates contextual data thus enhancing the learning process and provides 3D real models with additional explanations in order to help learners to broaden their knowledge. Several projects (Lai & Wang, 2012) have investigated the use of interactive AR in edutainment. The major benefits of applying mixed realities to educational environments consist in providing young users with a better understanding of concepts, the possibility of resizing and manipulating their augmented representations, personalized learning experience and the possibility to explore the virtual learning space at their own pace (self-directed learning). AR also supports acquisition of new skills such as problem solving, decision making, creativity and communication.

VR is an immersive multimedia 3D simulation of real life that supports interactivity with the created environment and enables sensorial experiences including virtual tastes, sights, smells, sounds and touches. VR has extremely wide applications across a whole range of disciplines, and the technology has reached a sufficient level of maturity to be applied in education in a wide area of topics, such as medicine (Izard, Ménèdez, & Palomera, 2017), mathematics and geometry (Moyer-Packenham et al., 2016) and engineering (Amirkhani & Nahvi, 2016). VR supports the development of the following skills: imagination, digital literacy, problem solving and risk taking.
VL is a highly interactive multimedia environment that brings learners into a computer-generated world that provides a simulation of the real world to be visually explored in a 3D environment and helps learners to achieve practical skills in an enjoyable and fun way. Numerous projects such as Virtual Engineering Science Learning Lab (VESLL, 2018) and Virtlab Virtual Laboratory (Virtlab, 2018) and research papers (Ghergulescu et al., 2018) propose to develop online interactive learning environments centred on a functional laboratory that supports collaborative problem solving and enhances students' practical skills and digital literacy.

“Interactive learning,” “active learning,” “knowledge building activities” and “21st century skills” are the key elements of edutainment. All the technology oriented teaching pedagogies discussed above support self-directed, interactive, edutainment environment and develop knowledge building skills required for the 21st century such as critical thinking, problem-solving, communication skills, creativity, imagination, interpersonal skills and lifelong learning attitudes. Knowledge building activities integrated in edutainment fosters 21st century skills that a person should have in order to become a lifelong learner, a person capable to adapt to the new technology developments that are emerging on the market at a growing speed.

Gibbons (2002) defined self-directed learning as “any increase in knowledge, skills, accomplishment, or personal development that an individual selects and brings about by his or her own efforts using any method in any circumstances at any time” (p. 2). Benefits of self-directed learning include: increased ownership of learning, fostering metacognition, increased workplace performance, nurture of appreciation for learning, increased academic achievement, motivation and self-efficacy (de Boer, Donker-Bergstra, Kostons, Korpershoek, & van derWerf, 2013; Guglielmino, 2013; Lavasani, Mirhosseini, Hejazi, & Davoodi, 2011).

This research paper investigates the effectiveness of combined Flipped Classroom (FC) and Problem-based Learning (PBL) teaching approach in a computer programming module delivered as a part of a skills conversion Computing course. Combined FC-PBL supports authentic learning in terms of authentic context, multiple perspectives through team work and collaboration. Both the educational and edutainment benefits are analyzed through pre-test, post-test assessments and survey. Knowledge building skills developed through combined FC-PBL are also investigated. Although FC and PBL pedagogies have been researched extensively, very few research papers have investigated a combined FC and PBL approach and have assessed the entertainment and educational benefits as this research aims to do.

**Literature review**

The use of technology in the formal and informal education can improve the learning experience and performance. Flipped classroom (Bradford, Muntean, & Pathak, 2014), interactive educational games (El Mawas et al., 2018a; El Mawas et al., 2018b; Ghergulescu & Muntean, 2010; Muntean, El Mawas, Bradford, & Pathak, 2018; Muntean, Andrews & Muntean, 2017) virtual labs (August et al., 2016; Bogusevski et al., 2018), enhanced learning experiences through augmented and virtual reality (Cai, Chang, Sun, Lin & Lee, 2017) are some of the technology enhanced pedagogies that have been applied in education. As this research investigates two self-directed technology based pedagogical approaches such as Flipped Classroom (FC) and Problem-based Learning (PBL) some recent research papers in this domain are discussed next. FC interconnects the home and classroom learning spaces making use of the latest technologies, while PBL enables more authentic learning activities. By freeing class time, the two pedagogical approaches enable the students to spend more time on practical activities, working on real problems and to connect what they learn in school with what they experience in their daily lives. The learning activities become more engaging, as the teachers become facilitators and assistants instead of instructors. Moreover, by using mobile and wireless technologies, seamless flipped learning enables even more authentic and fun learning activities as teachers can include field trips to leverage the diversity of experiences and facilitate the classroom discussion (Hwang, Lai, & Wang, 2015). Flipped classroom makes extensive use of educational videos which is characterised by an edutainment duality, that is to both educate and entertain (Cheng, Safont, Basu & Goebel, 2010). Todays’ students have become used to be entertained continuously, and previous research has showed that edutainment with humorous videos can increase students’ positive mood, and if the videos are congruent with the subject matter they help increase retention in the short and long term memory (Steffes & Duverger, 2012).

**Flipped classroom (FC)**

FC is defined as “a pedagogical approach in which direct instruction moves from the group learning space to the individual learning space, and the resulting group space is transformed into a dynamic, interactive learning
environment where the educator guides students as they apply concepts and engage creatively in the subject matter” (Flipped Learning Network, 2014, p. 1). The main value of flipped classroom is that by doing the instruction online before the class, it frees the class time for students to ask more questions, and for teachers to provide more feedback to individual students, to monitor their progress and to clarify misconceptions (Tucker, 2012). The flipped classroom pedagogy draws on various learning theories such as self-directed learning, active learning, student engagement and motivation, learning styles, cognitive load, and collaborative learning (Bishop & Verleger, 2013). Various research studies have investigated the benefits of applying flipped classroom, in different modules such as computer networking (Chen & Chen, 2014a; Chen, Wang, Kinshuk, & Chen, 2014b), mathematics (Bradford, Muntean, & Pathak, 2014; Love, Hodge, Grandgenett, & Swift, 2014), statistics (Touchton, 2015), programming (Karaca & Ocak, 2017), biology (Jensen, Kummer, & Godoy, 2015), or pharmaceutics (McLaughlin et al., 2014).

Chen et al. (2014b) have conducted a flipped classroom study over 18 weeks with 32 students enrolled in a networking course. The students spent around six hours per week, three hours reviewing videos and three hours attending a synchronous online class. For the synchronous class the authors followed previous recommendations from the literature such as to keep the design simple, to use a second screen for the lesson plan, to provide teacher-directed activities, etc. For evaluation purposes the authors used a survey, interview and system logs analysis. The results have shown that in general the students were satisfied with the new learning approach. However, the interviews revealed several barriers to the full adoption of the FLIPPED model, such as the students’ passive learning habits, and the busy schedules of part-time students that were not always able to properly review the study materials in advance of the synchronous class. The main limitations of the study were the low number of participants, and limited control of the experiment as the comparison with the previous two years was only done in terms of metrics such as grades, forum posts and attendance.

Jensen et al. (2015) have conducted a flipped classroom study for a biology course, where the hypothesis was that instructor facilitation is the main causal factor of improvements in student learning in the flipped model. The authors have tightly controlled all other factors, by exposing both non-flipped and flipped groups to the same active-learning instructional materials using the E-5 learning cycle (i.e., engage, explore, explain, elaborate, and evaluate). The first three phases of E-5 correspond to active content attainment and took place in class for the non-flipped group and online as homework for the flipped group. The last two phases of E-5 correspond to content application and took place online as homework for the non-flipped group, and in class for the flipped group. Both groups took 39 class assessments over the semester, out of which 13 assessments were also used in previous years (i.e., the original group used for additional control), while the rest were newly created assessments. In addition, the students had to take 3 unit exams and one final exam. The final exam consisted of 40 low level items, 40 high level items and 24 items to measure students’ reasoning. The main finding of the study was that flipped classroom does not have significant benefits over non-flipped classroom, as long as both follow an active learning approach. However, flipped classroom may be a good way to introduce active learning which has significant benefits over traditional passive learning.

Love et al. (2014) have conducted a flipped classroom study for a linear algebra maths course. The students were allocated to two groups based on their availability, without knowing that one group will be taught using the traditional approach and the other group using flipped classroom approach. The results analysis showed that the two groups achieved similar results in the final examination. However, the flipped group achieved a higher increase in grades for the second exam relative to the first exam, as well as for the third exam relative to the first exam. The analysis of a student perception survey showed that 74% of students had a positive attitude towards the flipped classroom approach. Moreover, the results showed that in-class board work helps students to become more comfortable with explaining to classmates and allowed them to develop deeper understanding.

Bradford et al. (2014) have applied flipped classroom to a first year course on introduction to mathematics for computing. Their approach was different, in the sense that only some of the topics were flipped, with all students attending both the flipped and the non-flipped topics. The lecture recordings were delivered through Moodle, and the course was structured so that the videos related to a topic would be viewed in sequence, by enabling watching subsequent videos. The evaluation results showed that the flipped classroom approach improved the students’ performance for continuous assessments, but did not impact on exam performance. Overall, the students perceived the flipped topics as being less difficult, and that students preferred flipped over non-flipped topics. However, the research findings were limited by the low number of survey responses and the lack of a control group.

Karaca and Ocak (2017) have applied the flipped classroom approach to a computer science course on algorithms and programming, in order to evaluate its impact on students’ cognitive load. The authors have used
interactive videos with embedded elements such as quiz questions, and links to other materials. The cognitive load was measured by asking the students to fill a 9-point scale at the end of each face-to-face lesson. The results showed that the students from the flipped group had lower cognitive load than the students from traditional face-to-face classroom. The main limitation of this study was that it was a bit narrow in scope by only looking to students’ cognitive load. Moreover, it did not use some baseline testing to check if the daytime and evening students are similar, and did not present other factors that may impact on their cognitive load.

McLaughlin et al. (2014) have applied flipped classroom into a course teaching the basics of pharmaceutics. The students were provided with online videos and reading resources, while the class time was dedicated to student-centred learning activities in the following order: open questions to assess students’ understanding, pair and share activities, student presentations and discussion, and individual or paired quiz. The instructor also used micro-lectures to reinforce and if needed to redirect students’ learning. The results showed that attendance and final exam results were higher for the flipped classroom as compared to previous non-flipped years. The survey revealed several positive effects, including that over 90% of students perceived the resources as being helpful, that flipped classroom improved their understanding, preparation for exams and their ability to apply the knowledge in the future.

Touchton (2015) has applied the flipped classroom technique to an advanced statistics course. The same videos and assessment project were used for both flipped and non-flipped groups, with the difference that the flipped group watched the videos at home. While the difference in the overall result was small, the flipped group performed considerably better on difficult topics such as: identification of methodology, presentation of results, and interpretation of findings. Moreover, the flipped classroom group gave higher ratings to the class and teacher, and felt they learned more than the traditional group.

A major limitation of all the described research studies is the reliance on grades and the possibility to introduce bias when marking different groups and/or analysing the data.

**Problem-based learning (PBL)**

PBL is a pedagogy in which students learn by solving real-world or open-ended problems. Savery (2015) defines PBL as “an instructional (and curricular) learner-centred approach that empowers learners to conduct research, integrate theory and practice, and apply knowledge and skills to develop a viable solution to a defined problem.” (p. 5). PBL is related to other pedagogies such as Project-based Learning, to whom it presents similarities such as the fact that both focus on open-ended questions, provide authentic applications of contents and skills, help build 21st century skills, emphasise student independence and enquiry, and are longer and more multifaceted than traditional lessons or assignments (Larmer, 2014). However, PBL activities tend to focus on smaller scale problems for which pre-existing solutions may exist, and are usually applied within the context of a single module. Various research studies have applied PBL in different modules such as computer science (Şendağ & Ferhan Odabaşı, 2009), software applications (Tsai, Lee, & Shen, 2013), earth science (Lawless & Brown, 2015), healthcare management and paramedic training (Beaumont, Savin-Baden, Conradi, & Poulton, 2014), or production of eBooks (Tsai, Shen, & Lu, 2015).

Şendağ and Ferhan Odabaşı (2009) have conducted an evaluation study with 40 participants that attended the Computer II course. The participating group was randomly split in half into experimental and control groups, after matched pairs were formed based on subjects’ prior content knowledge, prior critical thinking skills (CTS) scores, final grades from the Computer I course, Internet use hours per week, and gender. The experimental group was exposed to PBL activities involving ill-structured problem scenarios, while the control group was exposed to traditional teaching. For both groups the teaching was done through the Moodle platform. The results showed that both groups increased their knowledge between pre-test and post-test, but no group outperformed the other. However, the group exposed to PBL achieved statistically higher increase in critical thinking skills than the group exposed to traditional online teaching. The main limitation of the study was the low number of participants. Moreover, the analysis could have been expanded to compare the groups not only on the overall learning outcomes, but also on subsets of questions that measured different aspects such as comprehension and application.

Lawless and Brown (2015) have conducted an evaluation study with 535 urban and suburban middle school students for a science topic on water resources. The students were exposed to GlobalEd 2, a set of technology-mediated PBL simulations that “capitalise on the multidisciplinary nature of the social sciences as an expanded curricular space for students to learn and apply scientific literacies and concepts, while simultaneously also
enriching their understanding of the social sciences” (Lawless & Brown, 2015, p. 1). GE2 simulation consists of three phases lasting 14 weeks. Phase one requires the students to use text and web resources to research and identify the key scientific issues of concern, as well as how their assigned country’s culture, political system, geography and economy influence their science perspectives. Phase two requires students to work interactively within their class to refine their arguments and negotiate international agreements with the other “countries” sharing them online, in a secure asynchronous format similar to email. Phase three is designed to activate metacognition processes in students as they review what they learned and how they can apply it. The results analysis showed that GE2 PBL simulations had positive effects on students’ self-efficacy, their interest in science and their scientific writing.

Tsai et al. (2013) have conducted an evaluation study with 76 first year students from a vocational school that attended a one semester computing course called Packaged Software and Applications. The authors applied PBL and self-regulated learning (SRL) strategies. The students were divided in two experimental groups that attended a blended course and one control group that received traditional didactic lectures. The first experimental group was exposed to both PBL and SRL in the online environment, while the second experimental group was only exposed to PBL. The students were re-evaluated after 36 months to assess the effect of PBL and SRL applied in the first year on their long-term computing skills. The results showed that students exposed to online PBL and/or SRL had statistically significant better computing skills in month 36 than students exposed to traditional lectures.

Beaumont et al. (2014) conducted a study to evaluate how immersive virtual worlds could support PBL. The authors implemented two PBL scenarios in Second Life, for health care management and paramedic training. The first scenario was evaluated with 12 health-care professionals, while the second with ten first year paramedic students. The qualitative analysis of participants’ perceptions showed that Second Life PBL simulations provided a rich and engaging environment which enhanced authenticity of the scenarios, though there were issues of access and usability. The main limitation of the study was that it only conducted qualitative analysis, with no quantitative analysis of the other data collected, such as system logs being presented.

Tsai et al. (2015) have conducted an evaluation study with 144 elementary school students to compare three different approaches of teaching: problem-based learning with flipped classroom (FPBL), PBL and traditional teaching. The authors collected both quantitative and qualitative data, including interviews with students and teachers’ journal. The results showed that FPBL had statistically higher effect on improving students’ learning performance as compared to the other two teaching methods compared.

The main conclusion that can be drawn from the literature review is that previous research studies have mostly focused on evaluating the impact of a single technology-based pedagogical method, FC or PBL versus traditional teaching and learning. While the literature review identified a study that assessed the effects of combining FC with PBL (Tsai et al., 2015), that study was with elementary school students. To the best of our knowledge no previous research study has evaluated the impact of combining FC and PBL with university students. Another conclusion that can be drawn from the literature review was that FC or PBL do not always lead to higher knowledge gain as compared to traditional learning, thus more research studies are needed in order to thoroughly evaluate the benefits of the combined use of these pedagogical methods.

Therefore, the research study presented in this paper focuses on evaluating the effect of combining FC and PBL and applying them in a Software Development module delivered as part of a 3rd level skills conversion Computing course. The research study takes into account some of the limitations identified for previous studies in the literature, by considering both the educational and edutainment aspects into the evaluation.

**Case study description**

This section provides a description of the conducted research study that aimed to investigate the effectiveness of employing a combined FC-PBL approach to teaching a computer programming module delivered as part of a skills conversion Computing course. In addition, this study assesses the entertainment and educational benefits brought by the combined FC-PBL approach.

There are four main research questions identified and investigated in this research case study:
- Does the combined FC-PBL approach help students to improve their problem-solving skills?
- Do the students’ self-directed learning and knowledge building skills enhance?
- Do the students enjoy working in a collaborative environment as part of a team?
- Does the combined FC-PBL approach improve the students’ satisfaction?
PBL and FC teaching pedagogies were utilized in a face-to-face learning environment with students studying for a Software Development module delivered as part of Higher Diploma in Science in Computing degree, a one-year conversion course provided by National College of Ireland. 53 mature students (over 25 years old), 19.5% females and 80.5% males took part in the study. The case study was run over a 9 week duration learning session in three stages, each stage having a duration of 3 weeks. Figure 1 shows a summary of the three stages and the pedagogical approaches used during each stage. The first stage involved the use of a traditional approach, namely lectures and practical exercises, to teach programming concepts related to classes and objects. The second stage involved the use of FC pedagogy to teach programming concepts related to repetition statements. The combined PBL-FC pedagogy was used during the third stage for teaching arrays data structure related concepts. At the beginning of each stage the students took a pre-test that assessed the students’ knowledge on the topics to be covered during that stage. A few students may have had exposure to the topic outside of class, prior to commencing the course, and we wanted to ensure that we compare the results to an accurate baseline. At the end of each stage an in-class assessment (i.e., post-test) was run assessing the students’ knowledge level on the studied topics. The assessment components (i.e., pre-tests and post-tests) were developed by a panel of lecturers who have expertise in teaching software development-related modules at both undergraduate and postgraduate levels. Each lecturer proposed questions for the tests and the questions were cross-evaluated by the other members of the panel. The pre-tests and post-tests comprised of seven to ten questions each, depending on the complexity of the questions and the weight of a particular assessment component (i.e., post-test) towards the overall grade for the module. The types of questions used include the application of theoretical concepts to a real-world problem, understanding of code that the students have not seen before, and development of responses to well defined problems. A pair of pre-test and post-test was designed so that while each of the tests had different questions both tests covered the same topic. The questions were developed to evaluate the attainment of the expected learning outcomes. Overall, the tests were each assigned the same maximum score to enable reliable comparison across the different assessments.

The FC teaching approach required students to study basic programming concepts in advance of the class sessions by watching a set of short videos. Students were allowed to watch the videos as many times they wanted. Students were asked to do a short quiz after each watched video that was aimed to provide them feedback. The following programming concepts related to repetition statements were covered by the videos: for loop, while loop, do-while loop and nested loops. In return, classroom time was utilized for solving in-class practical exercises, and raising and answering questions. This approach aimed to capitalize on the time spent teaching in the classroom and to use the time to check the students’ knowledge on the topic in hand. A student who has watched the videos is more equipped to carry out in-class exercises which in return helps them to develop their conceptual understanding and programming skills. The lecturer’s role was to guide students as they applied concepts and engaged with the content. At the end of the 3 week duration FC based learning sessions, an in-class assessment (i.e., post-test) was run in order to assess the students’ knowledge level on the studied topics. The learning outcome was analyzed by comparing the pre-test and post-test results.

The combined FC-PBL pedagogy applied during the third stage was centered around group work. The students were presented with an open-ended real life problem that required knowledge on arrays data structure and they were required to problem solve by drawing on their own experiences. A video clip on the arrays concept was also provided in advanced of the class session and students were asked to watch it. A short quiz was run just after the visualization of the video. Authentic context based learning of the array concept and the use of the arrays in a software application was required in order to be able to identify a solution to the given open-ended real life problem provided in the class. The objective of this teaching strategy was to develop students’ critical thinking, reasoning, communication and team work skills. Students were allocated three weeks to collaborate by working in teams and to produce their findings and implement an application that solves the given problem. They were provided with the opportunity to ask questions during class and to try and work out the problem in their groups. Their activity and outcome was delivered as a project and assessed (i.e., post-test). During the teaching sessions, the role of the lecturer was to provide guidance and feedback.
The educational and edutainment benefits of using the combined FC-PBL approach in teaching programming concepts to mature students within an ICT skills conversion programme were investigated. Statistical analysis of pre-test and post-test assessments on concepts delivered through traditional, only FC, and combined FC-PBL teaching approaches was performed. The edutainment element was investigated through a questionnaire, provided at the end of the case study. The questionnaire consisted of 21 quantitative questions grouped in seven categories, namely problem solving skills, teamwork, self-directed learning, knowledge building skills, learning environment, satisfaction, and engagement as shown in Appendix A. The questions included in the questionnaire were designed following best practices and recommendations from literature on evaluating different approaches in teaching and learning. The questions that assessed the dimensions of engagement, enjoyment, and learning were designed following the recommendations from Feng, Chan, Brzezinski, and Nair (2008) and Mac Namara and Murphy (2017). Learning was assessed based on research work presented by Bourgonjon, Valcke, Soetaert, and Schellens (2010). Each question was answered on a 1 to 5 Likert scale (1 – never to 5 – always).

Results analysis

Statistical analysis of the data collected during the three stages of the study is presented and discussed in this section. The main aim of this research is to assess the effectiveness of employing a combined FC-PBL teaching approach. A different teaching approach was employed in each stage of the case study, namely traditional, FC, and combined FC-PBL. Pre-test and post-test knowledge assessment results for the three stages are analyzed and compared.

To be able to answer the research questions presented earlier (see Case Study Description section), in addition to using assessments (e.g., pre-test and post-test), quantitative data was gathered through a questionnaire. The questionnaire collected the students’ feedback regarding enjoyment, satisfaction, and engagement.

Analysis of learning outcomes and skills enhancement

An important evaluation of the effectiveness of using a combined FC-PBL approach to teaching is to analyse the learning outcome and to establish whether the students’ skills, and in particular the problem-solving skills have improved. Learning outcome is measured by using two assessments, namely a pre-test and a post-test. Figure 2 shows the assessment results for the combined FC-PBL approach, documented in terms of minimum, first quartile, median, third quartile, and maximum results. It is worth mentioning that an analysis of the pre-test assessment done before FC-PBL was applied showed that only 7.5% of students (i.e., 4 students) scored a mark of at least 70.0%, whereas 77.4% of students (i.e., 41 students) got a mark lower than 40.0% (i.e., 40.0% is the passing mark in the Irish education system) and 51.0% of students took a mark of 0. The post-test results analysis shows that 71.7% of students scored a mark of at least 70.0%, whereas only one single student scored less than 40.0%. The results of a paired t-test for dependent groups showed that the post-test results ($M = 81.0, SD = 18.4$) were statistically significant higher than the pre-test results ($M = 19.7, SD = 28.2$) at $\alpha = 0.001$ significance level ($t(52) = -15.73, p = .000$). The post-test results analysis also shows that the combined FC-PBL approach significantly improves the knowledge level.

The problem solving skills acquired through the traditional teaching approach, FC, and combined FC-PBL approach are assessed next to provide a quantitative measure of the effectiveness of the latter approach. Figure 3 shows the post-test results for the traditional teaching, FC-only, and combined FC-PBL approaches, namely minimum, first quartile, median, third quartile, and maximum results. Table 1 presents a summary of the post-test results for the three teaching approaches. The results of a paired t-test showed that the assessment results for combined FC-PBL ($M = 81.0, SD = 18.4$) were statistically significant higher than the assessment results for the traditional teaching approach ($M = 69.7, SD = 33.3$) at $\alpha = 0.01$ significance level ($t(52) = -2.78, p = .0075$). The results of a paired t-test showed that the assessment results for combined FC-PBL approach ($M = 81.0, SD = 18.4$) were statistically significant higher than the assessment results for FC only ($M = 64.0, SD = 28.5$) at $\alpha = 0.01$ significance level ($t(52) = -4.87, p = .000$). The results also show that when employing a combined FC-PBL approach, there was, on average, a 26.56% increase in the assessment results when compared to the assessment results achieved for FC only.
Figure 2. Pre and post assessment results for combined FC-PBL approach

Figure 3. Post-test results for traditional teaching, Flipped (FC), and combined FC-PBL

Table 1. Summary of assessment results for the three teaching approaches

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>First quartile (Q₁)</th>
<th>Median</th>
<th>Third quartile (Q₃)</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional</td>
<td>0</td>
<td>34.5</td>
<td>83.4</td>
<td>100</td>
<td>100</td>
<td>69.8</td>
<td>33.3</td>
</tr>
<tr>
<td>FC-only</td>
<td>12.5</td>
<td>40.7</td>
<td>62.6</td>
<td>100</td>
<td>100</td>
<td>64</td>
<td>28.5</td>
</tr>
<tr>
<td>FC-PBL</td>
<td>38.9</td>
<td>64.8</td>
<td>87.5</td>
<td>98</td>
<td>100</td>
<td>81</td>
<td>18.4</td>
</tr>
</tbody>
</table>

Table 2. Breakdown of assessment results by results category and teaching and learning approach

<table>
<thead>
<tr>
<th></th>
<th>Traditional</th>
<th>FC-only</th>
<th>FC-PBL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fail results – [0,40) [%]</td>
<td>28.3</td>
<td>24.5</td>
<td>1.9</td>
</tr>
<tr>
<td>Average results – [40,70) [%]</td>
<td>11.3</td>
<td>39.6</td>
<td>26.4</td>
</tr>
<tr>
<td>High results – [70, 100] [%]</td>
<td>60.4</td>
<td>35.8</td>
<td>71.7</td>
</tr>
</tbody>
</table>
The analysis of the assessment results demonstrates that by using a combined FC-PBL approach the knowledge acquired by students improves. In particular, the combined FC-PBL seems to prove effective in helping weaker students to improve their skills and knowledge due to different reasons such as sessions dedicated to provide guidelines and feedback, peer learning, and team work.

Edutainment analysis

The questionnaire served two main investigative purposes (a) student self-assessment of learning outcomes and skills gained, and (b) to provide a measure of the edutainment aspect of the combined FC-PBL approach. The learning outcomes and skills gained are quantified through questions classified in three categories, namely problem solving skills, self-directed learning, and knowledge building skills. Each of these categories contains 3 questions. Each question was answered on a 1 to 5 Likert scale (1 – never to 5 – always). Figure 4 presents the average responses provided for each question in the aforementioned categories. The results show that all except one question (i.e., Q12) were ranked with a score, on average, of at least 3.5. The aggregated feedback for problem-solving skills, self-directed learning, and knowledge building were rated, on average, at 3.7, 3.9, and 3.7 respectively. Many students considered that they developed their problem-solving skills and knowledge, hence 58.5%, 69.1%, and 57.7% of students rated each question in the problem-solving skills, self-directed learning, and knowledge building categories respectively with at least a 4. The questionnaire results show that the combined FC-PBL approach does contribute to the improvement of students’ problem-solving skills, self-directed learning, and knowledge building skills.

![Figure 4. Questions on problem-solving skills, self-directed learning, and knowledge building skills](image)

![Figure 5. Edutainment assessment: learning environment, satisfaction and engagement](image)

The edutainment component was measured in the questionnaire via three categories of questions that investigated the students’ engagement, and students’ satisfaction with the combined FC-PBL approach and learning environment. Figure 5 shows the average responses given by the students for the three categories. The students rated their satisfaction, on average, at 3.4 and their engagement, on average, at 3.7. The aggregated result for the learning environment was rated, on average, at 3.4. The students rated their motivation the highest, on average,
The research paper has investigated the educational and edutainment benefits of combined Flipped Classroom (FC) and Problem-based Learning (PBL) teaching approach in a computer programming module delivered as part of a skills conversion Computing course at National College of Ireland. The combined FC-PBL approach makes use of learning technologies and supports authentic learning in terms of authentic context, multiple perspectives through team work and collaboration.

A 9 week period, three-stage learning process that involved traditional teaching, FC-only, and combined FC-PBL approaches was applied. Pre-test, post-test, and team-based project assessments have been run as part of the course. The results show that when using a combined FC-PBL approach there was, on average, a 26.56% increase in the assessment results when compared to the assessment results achieved, for example, when only the FC approach has been employed. An analysis on the post-test assessment results breakdown for each of the three teaching approaches shows that combined FC-PBL is effective in helping weaker students to improve their skills and knowledge. For example, when the traditional and FC-only approaches were used 28.3% and 24.5% of students respectively scored a mark lower than 40% in the post-test assessment. In contrast, when the combined FC-PBL approach was employed only 1.9% of students scored a mark lower than 40%.

The edutainment element was investigated through a questionnaire, with answers on a 1 to 5 Likert scale (1 – never to 5 – always), which assessed engagement, learner satisfaction and whether the students have enjoyed the learning sessions. The results show that combining FC and PBL pedagogical strategies is effective in teaching programming concepts, supports knowledge acquisition though contextual study and real-life problem solving and provides an enjoyable learning experience. The students rated their satisfaction, on average, at 3.4 and their engagement, on average, at 3.7, and their motivation was rated the highest, on average, at 3.9. Furthermore, 52.7% and 61% of students provided a rate of at least 4 for the questions in the satisfaction and engagement categories respectively, therefore the combined FC-PBL approach does aid in the edutainment of mature students.

The analysis has also shown that the learning assessment results for the combined FC-PBL approach are statistically higher than the results for traditional and FC approaches. This is in line with the research finding of Tsai et al. (2015), which showed that combining problem-based learning with flipped classroom had statistically higher effect on learning performance as compared to PBL and traditional learning approaches. However, the authors did not compare the combined approach with FC. Moreover, their study was conducted with elementary school students, and to the best of our knowledge there is no other comparable research study that evaluated the combined FC-PBL approach with university students. On another side, the analysis showed that there is no statistical significant difference between the FC approach and the traditional approach in terms of learning assessment results. This is in line with the finding of some previous research studies conducted by Jensen et al. (2015) for a general biology course, and Love et al. (2014) for a linear algebra course.

The case study presented in this paper employed different pedagogical approaches to teach programming related concepts on a limited number of topics, namely classes and objects, repetition statements, and arrays. While it can be considered a limitation the fact that the combined FC-PBL has been used only on one topic, the results of this study do show that the blended approach is effective in teaching programming.

The research work presented in this paper is part of the NEWTON project, a large scale European project that designs, develops and deploys innovative solutions for Technology-Enhanced Learning involving delivery of state-of-the-art STEM content. A large-scale pilot that will deliver an entire programming (software development) course over 12 weeks period, through the NEWTELP platform is already scheduled. This pilot will be deployed in three universities across Europe and it will involve over 150 students. The course will use a blend of technology enhanced pedagogies such as game-based learning, flipped classroom and problem-based learning and gamification elements and it will use technology-enhanced learning materials. Educational content used in
the case study presented in this paper will be part of the NEWTON programming pilot. While research work on game-based learning has shown that the game helps the students to make sense of the abstract programming concepts, we would like to assess whether a game-based teaching approach enhances the edutainment aspect.

Acknowledgements

This research is supported by the NEWTON project (http://www.newtonproject.eu/) funded under the European Union’s Horizon 2020 Research and Innovation programme, Grant Agreement no. 688503.

References


Appendix A

Questionnaire

1. I give permission for the data gathered in this questionnaire to be used for a research study that assesses the effectiveness of PBL and Flipped classroom pedagogies in a classroom environment.
   [Yes] [No]

2. What is your gender?  [Female]  [Male]

3. What is your age range?
   - 18-21
   - 21-24
   - 25-34
   - 35-44
   - 45+

4. What course are you currently studying?

5. What category of student do you belong to?
   - School Leaver
   - Mature student
   - Other
   If other, please specify: ________________________________

Direction: Please select and rate the following questions using the following scales:

<table>
<thead>
<tr>
<th>5 – Always</th>
<th>4 – Often</th>
<th>3 – Sometimes</th>
<th>2 – rarely</th>
<th>1 – never</th>
</tr>
</thead>
</table>

6. Problem Solving Skills
   Q1. I can apply a variety of problem-solving approaches
   Q2. The project has helped me to develop my problem-solving skills
   Q3. I feel confident about tackling unfamiliar problems

7. Team Work
   Q4. I enjoy working as part of a team
   Q5. I am willing to forego personal goals for the benefit of the group
   Q6. I am able to express disagreement or disappointment directly

8. Self-Directed Learning
   Q7. I am comfortable using variety of means to explain programming concepts to others
   Q8. I am willing to draw on my peers as resources
   Q9. I am willing to consider a wide range of learning resources

9. Knowledge building skills
   Q10. I am able to identify gaps in my knowledge in relation to the learning activity
   Q11. I can easily embed new knowledge by applying it to solve problems
   Q12. I can easily link theory to practice

9. Learning Environment
   Q13. The learning environment is suitable for me to participate in PBL
   Q14. The Flipped Classroom (videos) is more engaging than traditional classroom teaching

10. Satisfaction
    Q15. I would like PBL to be used in other modules on the programme
    Q16. I feel that PBL enhanced my experience in this module
    Q17. I like watching the lessons on video
    Q18. I like taking my quizzes online using Moodle
    Q19. Short flipped learning videos are more effective than traditional face-to-face lectures
## II. Engagement

<table>
<thead>
<tr>
<th>Question</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q20. I felt motivated to complete the PBL Project</td>
<td></td>
</tr>
<tr>
<td>Q21. I am satisfied with the work I did on the PBL project</td>
<td></td>
</tr>
</tbody>
</table>
iAbstract: Game-driven Keyword Auction and Summarization for Academic Reading

Hercy N. H. Cheng¹, Calvin C. Y. Liao²* and Wan-Chen Chang³

¹National Engineering Research Center for E-Learning, Central China Normal University, China // ²College of Nursing, National Taipei University of Nursing and Health Sciences, Taiwan, R.O.C // ³Department of Human Development and Family Studies, National Taiwan Normal University, Taiwan, R.O.C // hercycheng.tw@gmail.com // CalvinCYLiao@gmail.com // altheawcc@ntnu.edu.tw

*Corresponding author

ABSTRACT

Graduate students usually lack sufficient ability to read academic papers effectively and efficiently. For facilitating their reading comprehension, this study adopts summarization as a main reading strategy under the scaffold of keyword evaluation. Furthermore, this study attempts to transform summarization into a group-based educational game by incorporating keyword auction mechanisms. The research purpose is to investigate how first-year graduate students react to such an educational game when they are required to summarize a section of a real journal paper. The results indicate that they improve the completeness of the summaries without decreasing their conciseness after the game. Besides, the analysis of their prior summaries suggests that graduate students may lack the ability of structure analysis, idea integration and argumentation. Additionally, the results also show that there are positive relationships among the students’ evaluation, experts’ evaluation, auction behaviors and frequencies in summaries of the keywords, suggesting that students’ keyword evaluation may change their decision-making behaviors and the quality of their final summaries.

Keywords

Academic reading, Summarization, Game-based learning, Auction mechanism

Introduction

While undergraduate education usually focuses on teaching well-organized knowledge, graduate education allows students to explore emerging knowledge. Because graduate students have to conduct real and innovative research, they should learn how to do research in an authentic academic context so that they could develop their academic reading and writing abilities. These abilities involve tacit knowledge, which is usually hard to learn by lectures. This is why graduate education usually adopts apprenticeship, where students can learn from their advisors and academic papers. Furthermore, they read literature, discuss with their advisors, conduct a series of researches, and write academic papers. Through the process, they find their own purposes and meanings. The authenticity is necessary for graduate students to absorb the knowledge and know-how of doing research (Herrington, Reeves, & Oliver, 2014).

Brown, Collins, and Duguid (1989) described the authenticity as meaningful, purposeful, but ordinary practices of the culture. In general, the researchers of educational technology realized the authenticity in two ways: establishing authentic environments and providing authentic problems (see Herrington & Oliver, 2000). In authentic learning environments, students are allowed to interact with real or simulated people and artifacts. For doing so, researchers adopted the technology of virtual or augmented reality so that students may learn in real situations. For example, Huang, Yang, Chiang, and Su (2016) allowed students to use their own mobile phones to learn English vocabulary in outdoor learning sites. On the other hand, authentic problems require students to enquiry and solve contextual learning tasks, so that they may apply and reflect the knowledge in real life (Maina, 2004). For example, Hwang, Chiu, and Chen (2015) designed a contextual educational game, which could situate students in a series of inquiry quests in social studies. In a sense, game provided a reasonable context so that students may feel more engaged in solving complicated learning tasks.

Although authentic problems usually preserve the complexity of the real-life setting, these problems are relevant to students. In the process of problem solving, the students may practice their expertise and share with others. Like in graduate education, for doing appropriate research and writing academic papers, a fundamental capability is to read sufficient related papers. Furthermore, graduate students have to locate, organize, and explain important knowledge in their own way and for their own research purposes. However, several thousand Chinese students in a university reported their difficulties of getting the meaning of the article and determining the main ideas (Evan & Green, 2007), because novice academic readers usually lack prior knowledge to distinguish the contributions of the authors from previous works. They also do not have a specific goal to locate relevant content in the beginning. In this vein, grasping the gist and identify useful information become the first challenge of the first-year graduate students.
Among various reading strategies, summary writing may well facilitate students’ reading comprehension by synthesizing important ideas in papers (Carroll, 2008), while keywords evaluation may help students grasp the main ideas more easily (Chou, 2012). Although researchers have been developed many effective systems for learning how to read better, the effects seem decreased over time, especially for low-ability students (Magliano et al., 2005). After using these strategies for a while, the students may easily feel unmotivated. For these reasons, this study aims to integrate summarization with keyword evaluation in order to design a group-based game, iAbstract, for graduate students to summarize academic papers. In a sense, this study gamifies the process of academic summarization with several techniques of text analysis. Furthermore, this study adopts Chinese word segmentation technique to extract and weight keywords and latent semantic analysis to calculate the scores of summaries for supporting the gameplay. However, the educational technology researchers’ endeavor to gamify reading strategies is still in an early stage, because the reading and writing abilities could not be easily evaluated if researchers would like to avoid explicit assessment in games. Recently, as text analysis continues to a mature technology, the research on game-based reading and writing become more feasible. At this stage, it is expected that this study may enrich our understanding of the relationship between students’ game behaviors and learning outcomes. Besides, this study also concerns how graduate students compose their academic summaries in such a game-based learning environment.

Related works

Academic reading and summarization

Recent studies on academic reading focus on students’ learning experience more than learning performance (e.g., Liu, Chang, Yang, & Sun, 2011). Furthermore, some studies adopted reading-writing connection activities to improve students’ academic reading performance. For example, Shih (1992) classified academic reading lessons as three main phases: pre-reading, guided reading and post-reading exercises. Among the three phases, post-reading exercises, such as note taking and summarizing, are useful strategies for reducing details and organizing important ideas in papers. Baker and Boonkit (2004) found that the frequently used strategies for academic reading and writing were general problem-solving strategies, such as reading for the main idea of passages, connecting prior knowledge and new one during writing, etc. Besides, Kim and Kim (2017) found an integrated reading-to-write task could help graduate students’ capabilities of getting main ideas from the academic articles. In other words, writing activities after academic reading could help students’ learning performance.

Summarization, or summary writing, is a post-reading activity. Previous studies have shown that students who participated in summary writing have better reading comprehension than those who did not (Carroll, 2008). This is due to the fact that summarization not only requires students to read the article carefully, but also need them to understand its content more deeply. Furthermore, for writing summaries, students have to identify the important elements of the article and integrate the past knowledge with the new one (Wade-Stein & Kintsch, 2004). In other words, in order to write better summarization, students should comprehend the article first. Besides, summarization may also guide students to focus on the important ideas of the article and to build the relationships among them. By doing so, when summarizing an article, students can make better deletion, retention and modification of the content. In this process, students can reorganize the ideas and then gain better comprehension.

In general, the theories of summarization suggested that students should carry out the process of material deletion, word composition, and topic sentence invention (McNeil & Donant, 1982). In material deletion, students have to evaluate which information is unnecessary and redundant for the gist of articles. In other words, they determine the importance of the materials in articles. In word composition, students are encouraged to select and adopt a series of short words to replace the remained sentences. For example, Chou (2012) adopted keywords and questions. More specifically, after reading an article, students had to write three keywords that they thought important, and to answer 5W1H questions (i.e., what, why, where, when, who, and how) for writing summaries. By doing so, students can highlight what they should pay attention. Later, in topic sentence invention, students are required to write topic sentences to contextualize those words for reorganizing the gist of the articles.

Computer-based feedback for summarization

Because summarization is so important for reading comprehension, many researchers have designed various learning activities and scaffolds, so that students can write high-quality summaries (Konuk, Ören, Benzer,
Sefer, 2016). Although it is critical for students to practice summarization, grading students’ summarization or providing or feedback are still time-consuming and laborious tasks for teachers. Therefore, many studies adopted a variety of computer-based techniques to overcome this problem (Kintsch et al., 2000; Wade-Stein and Kintsch, 2004; Landauer, Lochbaum, & Dooley, 2009; Sung, Liao, Chang, Chen, & Chang, 2016). For example, He, Hui, and Quan (2009) reviewed several advanced techniques for supporting computer-based feedback of summarization, such as latent semantic analysis (LSA; Kintsch et al., 2000). The LSA has been adopted to design computer-based feedbacks, so that teachers may carry out the grading of summaries more efficiently.

For example, Kintsch et al. (2000) designed a first prototype of a computer-based feedback system, called State the Essence. In the system, they adopted the LSA technique to calculate the relationships between students’ summaries and source texts. Besides, the system also provided students with feedback for improving summaries, such as spelling, length, and overall content coverage. Wade-Stein and Kintsch (2004) then developed a Summary Street system to assess students’ summaries and provide feedback. Furthermore, they indicated that the students with feedback might spend more time on writing their summaries. Besides, they also found that the feedback demonstrated a lasting effect on students’ ability to summarize articles. In addition to providing feedback, researchers also have begun to integrate different scaffoldings in the past decade. For example, Landauer, Lochbaum and Dooley (2009) further developed WriteToLearn, a computer-supported writing system for reading comprehension. Moreover, the system provided students with tutorial feedbacks according to their traits of writing. In recent year, Sung et al. (2016) integrated concept maps to provide students with semantic and conceptual scaffolds. Their finding showed that the proposed system could improve the students’ summary writing skills.

Digital game-based reading activities

For engaging students in learning strategies for reading, many research designed various virtual characters. For example, McNamara, Levinstein, and Boonthum (2004) designed an iSTART system for training students’ reading strategies. In the system, more specifically, three virtual characters, including an intelligent tutor and two virtual students, demonstrated reading strategies for learners. These learners may learn how to paraphrase, elaborate and explain the meaning of selected articles (McNamara, 2009). Besides, Park and Kim (2015) developed a Virtual Tutee System to enhance college students’ academic reading engagement. In the system, after reading assigned articles, the college students were required to teach virtual tutees by providing lecture notes and answering the tutees’ questions.

The concept of virtual characters stimulates the research of game-based learning. Furthermore, some researchers started to incorporate game-based interaction among students. For example, Dempsey, Jackson, and McNamara (2010) integrated game characteristics into the iSTART system and developed MiBoard, a multiplayer online board game. In this game, students played the role of an explainer in turn. The explainer was required to use randomly selected strategies to explain a sentence from an article, while the other players guessed which one strategy the explainer used. Given that there was not much research on digital game-based reading, the MiBoard system inspired our design. Furthermore, game mechanisms are used to provide students with clear and positive goals for writing better summaries, while group-based activities may facilitate interactions among players.

Game design

Design framework

As McNeil and Donant (1982) suggested, a tutorial approach to summarization includes material deletion, word composition, and topic sentence invention. According to this approach, this study adopts a design framework, illustrated in Figure 1(a). In this framework, for writing summaries, students are required to evaluate the importance of information in articles, select appropriate keywords, and composing topic sentences. Following this framework, the study designed an educational game by integrating summarization and keyword auction to support students to participate in summary writing with pleasure. Furthermore, the summarization game requires every student to read an article, acquire important keywords, and then use the keywords to compose a summary.

An auction game mechanism is applied to the process of keyword selection. More specifically, students attempt to offer higher bids for high-valued keywords than the others. Meanwhile, they also have to control their costs to avoid offering too much bids. For doing so, they have to mentally evaluate the importance of keywords and
make correct decisions on bidding. In a sense, the auction mechanism is to compare students’ evaluation of keywords with others.

The game consists of four phases: reading, auction, summarization, and scoring.

1. Reading: Students use their own personal devices to read an academic article within a limited time. Meanwhile, the students are also required to evaluate selected keywords. For quantifying the importance of the keywords, this study adopted the Chinese word segmentation technique of Nature Language Processing and Information Retrieval Platform (NLPIR) of Chinese Academy of Science to produce a set of keywords with weights, which were calculated according to their frequencies and the similarities to the semantic space of the article.

2. Auction: Students are required to bid a reasonable price for important keywords. Furthermore, they plan auction on their personal devices, and then perform the action face to face. The auction mechanism will be introduced in detail in the next section. In the end of the auction, their devices show the selected keywords with different scores, which are assigned secretly depending on the weights of keywords in the article.

3. Summarization: Students are required to write summaries with keywords. Furthermore, they have to use the purchased keywords, but they are not allowed to use the keywords they do not buy. This rule is designed to make students select keywords more seriously in the previous stage. If they did not follow the rule, they would be asked to re-write the summaries.

4. Scoring: Finally, the game score is consisted of keywords importance, costs, and summary qualities. The keyword importance is equal to the scores of the purchased keywords (ranging 2 to 5 points per keyword). The costs are evaluated by students’ remaining money (1 points for every three remaining coins). The summary qualities are automatically calculated by the technique of latent semantic analysis for the gameplay. More specifically, the similarity of every sentence in their summaries to the article was calculated and transferred into 15 points at most. Therefore, for winning this game, students have to write better summaries, purchase higher-value keywords and pay lower prices. Finally, the player with the highest score wins the game.

### Auction mechanism

In the auction phase, the students are given 30 game coins and start to plan their auction. The auction planning allows students to consider all keywords in advance at the same time, instead of bidding one keyword at a time. Such design not only requires the students to have a whole picture, but also facilitates them to compare the importance of these keywords, lessening the influence of other factors, such as the cost. Furthermore, they are given a set of bidding cards, including one A class card, two B class cards, two C class cards, and three D class cards, as shown in Figure 1(b). The A card allows the students to acquire the keyword they wanted without any competition, but this card may increase its price by four coins. The B cards, which may increase the prices of the keywords by three coins, allow the students to acquire the keywords only when the other players do not use any A cards to bid. Similarly, the C bidding cards may increase the prices by two coins and allows them to acquire the keywords only when the others do not use any A or B bidding cards. The D cards are used to specify that the students do not want the keywords, but the cards may still increase the prices by one coin. Such design makes sure that students can acquire at least one keyword, avoiding the problem of failed auction.
After all students plan their auction secretly, the system calculates the prices of the keywords according to all bidding cards. After the prices are displayed, every highest bidder has to decide whether he/she wants to buy the keywords at the price. If he/she wants the keyword, then he/she has to pay the price. If not, the bidders with next class cards have the chance to decide whether to buy it. Furthermore, although students offer prices at the same time, the prices are varied and aggregated from all students’ evaluation. In other words, if all students wanted the same keyword, the price becomes higher. In a sense, the prices may reflect students’ evaluation of importance, which may allow them to compare their own evaluation with each other.

As mentioned earlier, the keywords are assigned with different scores, ranging from 2 points to 5 points. For preventing them from the influence of the scores, the scores are revealed in the last phase. Besides, in order to facilitate them to buy as more important keywords as possible, those coins will be devalued if they are not spent. Students have to evaluate whether the prices are worth enough. In an ideal case, students, who are able to exchange coins to keywords with equivalent values, have to use these keywords to compose better writings for winning. Otherwise, if any other players underestimate a keyword, they may have a chance to score more. Alternatively, if other players overestimate it, they may decide not to buy it and save the money to buy other keywords.

Method

Research questions

This study attempts to investigate how students compose their summaries and how they make evaluations in such an educational game. Furthermore, this study examines the relationships among students’ comprehension of the articles and auction behaviours as well as how these factors influence their summary writing. More specifically, there are two research questions: (1) How does the game change students’ summaries? (2) What are the relationships among students’ evaluation of keywords, summaries, and auction behaviours?

Reading material

Article selection

This study selected the Introduction section of a Chinese journal paper, entitled “The study of flow experience in digital game-based learning” (Chen, Yang, & Wen, 2016), as the learning material. This article was published at a Chinese Social Science Citation Indexed journal “Digital Education Research.” The article aimed to examine the influences of students’ academic ability and social interaction tendency on their flow experience in digital game-based learning. Before the study, the participants were asked about their knowledge of flow experience. The main ideas of this article are listed in Table 1. Each main idea should have an argument with explanation.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Main Ideas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research motivation</td>
<td>Argument: The reason for studying flow experience</td>
</tr>
<tr>
<td></td>
<td>Explanation: Definition of flow experience</td>
</tr>
<tr>
<td>Hypothesis of variable 1</td>
<td>Argument: The hypothesis of academic ability</td>
</tr>
<tr>
<td></td>
<td>Explanation: The perspectives of flow theory</td>
</tr>
<tr>
<td>Hypothesis of variable 2</td>
<td>Argument: The hypothesis of social tendency</td>
</tr>
<tr>
<td></td>
<td>Explanation: The perspectives of motivation theory</td>
</tr>
<tr>
<td>Hypothesis of interaction</td>
<td>Argument: The hypothesis of interactions between the two variables</td>
</tr>
<tr>
<td></td>
<td>Explanation: The perspectives of cognitive load</td>
</tr>
<tr>
<td>Research objective</td>
<td>Argument: The objective for studying the influence of the two variables and their interaction on flow experiences</td>
</tr>
<tr>
<td></td>
<td>Explanation: The absence of previous studies</td>
</tr>
</tbody>
</table>

Keywords extraction

This study extracted and selected eight keywords from the article for the game. More specifically, this study adopted the NLPIR Platform of Chinese Academy of Science to produce the top 50 important keywords of the article. As sown in Table 2, the researchers selected two keywords (“games” and “digital”) from top 10 keywords...
and assigned them as 5 game points. Similarly, the keywords with 4 and 3 points were selected from top 11–20 keywords and top 21–50 keywords, respectively. Finally, the researchers selected additional two keywords that were not included in top 50 keywords and assigned them as 2 points.

<table>
<thead>
<tr>
<th>Rank of weights</th>
<th>Selected keywords</th>
<th>Meaning in the article</th>
<th>Game scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 1-10</td>
<td>Games</td>
<td>Research context: digital games, game-based learning (environments)</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Digital</td>
<td>Research context: digital games, digital game-based learning</td>
<td>5</td>
</tr>
<tr>
<td>Top 11-20</td>
<td>Learning</td>
<td>Research context: learning in games</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Load</td>
<td>Backup arguments for research purposes: the perspectives of cognitive load</td>
<td>4</td>
</tr>
<tr>
<td>Top 21-50</td>
<td>Teaching</td>
<td>Research context: Instruction activities, applying game-based learning to instruction</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Interaction</td>
<td>Research purposes: statistical interactions between two independent variables</td>
<td>3</td>
</tr>
<tr>
<td>Not in top 50</td>
<td>Internal</td>
<td>Descriptions of cognitive load</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Hot issues</td>
<td>Descriptions of game-based learning</td>
<td>2</td>
</tr>
</tbody>
</table>

**Procedure**

The study involved 32 first-year graduate students (8 males and 24 females). Each session of the experiment recruited four students. Before a game, the students were required to answer questions about their prior knowledge of the topic. There were 29 (90.62%) students who did not correctly explain the knowledge, while only 3 students who were fully or partially correct about it. As a coordinator, a researcher then introduced the game rules and started the activity. In the reading phase, the students had to read the article in 10 minutes. Besides, they also had to sort the selected eight keywords according to their perceived importance to the article and write down prior summaries on their devices. The prior summaries were used to help them preliminarily draft the main ideas of this paper. Meanwhile, the researcher could collect their initial thoughts about this paper. In the auction phase, they planned their auction secretly and then decided whether to buy keywords. Highest bidders paid the prices and used the keywords to compose final summaries, which must included five sentences. After the scoring phase, finally, they were also required to individually explain their behaviors in the game.

**Data collection and analysis**

During the game, students’ summaries, evaluation of keywords, and behaviors of auction are all collected. Students’ summaries were denoted by the format like “S11-final.” The “S11” represented the identification number of one student, while the “final” represented that it was a final summaries rather than a prior one. Besides, two independent raters evaluated the quality of students’ prior and final summaries according to Table 3. As mentioned earlier, article was consisted of five main ideas, starting from research motivation, followed by three hypotheses and concluded with research objective. As shown in Table 3, a complete main idea should include a concrete argument with detailed explanation. Kappa statistic was adopted as a measure of agreement between the raters. The Kappa values were .855, suggesting a consistent agreement.

**Table 3. The criteria of each main idea in summaries**

<table>
<thead>
<tr>
<th>Quality</th>
<th>Definition</th>
<th>Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete</td>
<td>A concrete argument with detailed explanation</td>
<td>4</td>
</tr>
<tr>
<td>Flawed</td>
<td>Only an argument without detailed explanation</td>
<td>3</td>
</tr>
<tr>
<td>Incomplete</td>
<td>Only explanation without a concrete argument</td>
<td>2</td>
</tr>
<tr>
<td>Unrelated</td>
<td>Statements unrelated to the main idea</td>
<td>1</td>
</tr>
<tr>
<td>None or Wrong</td>
<td>Not mentioned any statements or wrong statements</td>
<td>0</td>
</tr>
</tbody>
</table>

Besides, this study also adopted terse values to evaluate the terseness of students’ writing, such as note taking or summary writing (Chang, & Ku, 2015). The terse values could be regarded as an informational density, indicating whether the students could sufficiently elaborate the main ideas by composing brief words. The terse values were calculated by the number of main ideas divided by the number of words (i.e., Chinese characters, in
this paper). Furthermore, if a student could write more ideas with fewer words in a summary, his/her terse value would be higher.

Results and discussion

The change of summaries

Table 4 shows the changes of the students’ summaries after the game. Statistical tests showed that the students significantly improved their summaries in terms of the numbers of main ideas and words. In other words, after the game, the students became able to write more main ideas in longer summaries. However, the change of the terse value was not significant, suggesting that the conciseness of important information remain the same. The stable terse values also implied that these graduate students did not increase words to write unimportant information. Instead, after the game, they were able to notice more complete main ideas. This was probably due to the reason that keywords may highlight what they should pay attention to and reorganizes the gist (Chou, 2012).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Prior summaries</th>
<th>Final summaries</th>
<th>t</th>
<th>SE</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td># of main ideas</td>
<td>1.406 (0.920)</td>
<td>1.836 (0.856)</td>
<td>2.396*</td>
<td>0.179</td>
<td>.023</td>
</tr>
<tr>
<td># of words</td>
<td>85.625 (44.661)</td>
<td>128.906 (35.931)</td>
<td>5.133***</td>
<td>8.432</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Terse value</td>
<td>0.015 (0.009)</td>
<td>0.014 (0.006)</td>
<td>0.711</td>
<td>0.002</td>
<td>.482</td>
</tr>
</tbody>
</table>

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

Table 5 further shows the completeness of each main idea before and after the game. Several t tests showed that the students significantly improved the completeness of their ideas about the research motivation and objective. However, their ideas about the hypotheses of independent variables and interaction were not significantly changed. It could be noted that, before the game, these graduate students did not particularly pay attention to research motivation and objective in the section of introduction. Instead, they focused on the explanation of each hypothesis and overlooked the whole picture. By evaluating the keywords in the game, the students could better organize their summaries.

<table>
<thead>
<tr>
<th>Main ideas</th>
<th>Prior summaries</th>
<th>Final summaries</th>
<th>t</th>
<th>SE</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research motivation</td>
<td>0.938 (0.716)</td>
<td>1.281 (0.924)</td>
<td>2.156*</td>
<td>0.159</td>
<td>.039</td>
</tr>
<tr>
<td>Hypothesis A</td>
<td>1.813 (1.491)</td>
<td>2.156 (1.370)</td>
<td>1.187</td>
<td>0.290</td>
<td>.244</td>
</tr>
<tr>
<td>Hypothesis B</td>
<td>1.594 (1.521)</td>
<td>1.594 (1.521)</td>
<td>0.000</td>
<td>0.269</td>
<td>1.000</td>
</tr>
<tr>
<td>Hypothesis A×B</td>
<td>0.781 (0.792)</td>
<td>1.000 (1.107)</td>
<td>1.070</td>
<td>0.204</td>
<td>.293</td>
</tr>
<tr>
<td>Research objective</td>
<td>0.500 (0.950)</td>
<td>1.313 (1.355)</td>
<td>4.104***</td>
<td>0.198</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

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

These findings were basically consistent with the survey of Evan and Green (2007), in which Chinese students found their difficulties in determining the main ideas of academic articles. This study further revealed the reading problems of the first-year Chinese graduate students. Specifically, when the students read the section of the introduction in an academic paper, they usually overlooked the structure of the article. They focused too much on relatively unimportant details. As a result, they could not form a systematic summary to fairly describe the content. In this vein, they probably lacked the ability of structure analysis and idea integration. Before reading, students should look over the text first and note the text structure (Duke & Pearson, 2002). One of the solutions was the scaffold of concept maps (e.g., Sung et al., 2016). Such graphical representation might help students organize better summaries. On the other hand, although the students improved their descriptions of research motivation and objectives in this study, there was still much room for increasing the completeness of their summaries. For each main idea, most of the students put too much emphasis on describing the explanations and forgot the arguments. This could result from lacking the ability of argumentation. By scaffolding students to make arguments, they should be able to construct more logical summaries.
The relationships among the evaluation and usage of keywords

Table 6 shows the spearman correlation coefficients among the evaluation and usage of keywords. As shown in the table, the students’ evaluation of keywords had significantly medium positive correlation with the evaluation of experts, suggesting that these graduate students in general were able to distinguish the importance of the keywords. Besides, the frequencies of the keywords in the students’ prior summaries were also positively related to experts’ evaluation. Although the frequencies of the keywords in prior summaries were associated with students’ evaluation, the correlation was low. The reason may be that when the students composed summaries at the first time, they had to focus more on the original article rather than considering the importance of the keywords.

<table>
<thead>
<tr>
<th>Table 6. The Spearman correlation coefficients among the evaluation and usage of keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spearman correlation coefficients</td>
</tr>
<tr>
<td>----------------------------------</td>
</tr>
<tr>
<td>(1) Expert evaluation</td>
</tr>
<tr>
<td>(2) Student evaluation</td>
</tr>
<tr>
<td>(3) Frequencies in prior summaries</td>
</tr>
<tr>
<td>(4) Bids offered</td>
</tr>
<tr>
<td>(5) Purchase</td>
</tr>
<tr>
<td>(6) Frequencies in final summaries</td>
</tr>
</tbody>
</table>

Note. “p < .01,” “**p < .001.”

After the activity, the frequencies of the keywords in their final summaries had a lower correlation with experts’ evaluation, and had a higher correlation with their own evaluation. The result suggested that they later became freer to use these keywords in their summaries. Interestingly, the behaviors in the game were also associated with these factors. Furthermore, there was a significant moderate positive correlation between students’ evaluation and bids offered in the games, while there was a low correlation between their bids and experts’ evaluation. The result suggested that they more likely offered high bids to their important keywords.

However, the behaviors of purchase were not necessarily related to these factors. Furthermore, the behaviors of purchase were mainly associated with students’ evaluation and bids, but not associated with experts’ evaluation or the usage in their prior summaries. This was probably because their purchase behaviors were constrained by the game rules. They could only make decisions to buy the items of their highest bids. Similarly, the frequencies of the keywords in the final summaries were highly related to what they had purchased.

On the other hand, students’ explanation of their behaviors may reveal their opinions on this game. Some students indicated the reason why they liked this game. For example, S42 mentioned that, “there was high uncertainty in the auction, in which the players not only had to tell the importance of the words, but also had to guess the bids offered by the opponents.” Similarly, S13 said, when bidding keywords, she “had to speculate the mental activities of other players, requiring players’ multiple abilities.” Another student S81 expressed his preference of strategy games, because “the mental competition may increase the playfulness.” However, some students did not like the type of auction games. For example, S11 said that, she “had to buy her wanted keywords at a very high price,” which was “less relevant to reading.” These opinions suggested that the playfulness of the game was somehow related to its decision making process. Furthermore, when the students made the decisions to offer bids, they had to consider both whether the keywords had high values, and whether other players would offer high bids. The complexity of situation analysis required students’ reasoning and speculation, likely resulted in playful thinking process. Unfortunately, the complexity could also be a two-edged sword. Some students who were not good at mental evaluation could feel frustrated and anxious. How to balance the complexity of games and playfulness will be an interesting issue to explore in the future.

Among all keywords, there were three keywords of which students’ evaluation were significantly positively related to the bids. As shown in Table 7, they were “games,” “load,” and “interaction.” The results suggested that when the students offered bids to the three keywords, they considered their importance to the article. Conversely, when offering bids to the other keywords, they considered other factors rather than the importance. The students reported that when they bidding keywords, they also considered “whether the keywords could be replaceable in addition to their importance (S11),” and “to spent as few as possible (S62).” If they could not purchase their wanted keywords, they may use “synonyms to replace the keywords (S42)” or “similar sentences to express the same meanings (S52).”

Previous researchers required students to generate keywords before summary writing (Chou, 2012), while this study allowed students to evaluate keywords and make decisions. Both methods may facilitate students to
improve their summaries. This may be due to the fact that composing short words to replace main ideas in articles was crucial to summary writing (see Duke, & Pearson, 2002). Furthermore, when summarizing articles, students were suggested to delete unnecessary information first, abbreviate main ideas as keywords, and finally invent topic sentence. Taking the advantage of games, this study provides students with opportunities to evaluate the importance of keywords instead of generating them. By doing so, they could mentally construct the relationship between keywords and summaries. By comparing the collective evaluation with their mental evaluation, the students had to make tough decisions. The results suggested that students’ mental evaluation was correlated with their game behaviors and the usage of keywords in the summaries.

<table>
<thead>
<tr>
<th>Keywords</th>
<th>Students’ evaluation (1~4)</th>
<th>Students’ bids (1~4)</th>
<th>Spearman correlation coefficients</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Games</td>
<td>3.22 (0.66)</td>
<td>3.03 (0.95)</td>
<td>0.564**</td>
<td>0.001</td>
</tr>
<tr>
<td>Digital</td>
<td>3.04 (0.78)</td>
<td>2.16 (0.94)</td>
<td>0.089</td>
<td>0.629</td>
</tr>
<tr>
<td>Learning</td>
<td>3.16 (0.69)</td>
<td>2.44 (0.90)</td>
<td>0.052</td>
<td>0.777</td>
</tr>
<tr>
<td>Load</td>
<td>2.31 (0.81)</td>
<td>2.03 (0.98)</td>
<td>0.411*</td>
<td>0.019</td>
</tr>
<tr>
<td>Teaching</td>
<td>2.25 (0.92)</td>
<td>1.91 (0.98)</td>
<td>0.282</td>
<td>0.120</td>
</tr>
<tr>
<td>Interaction</td>
<td>2.55 (0.76)</td>
<td>2.31 (0.98)</td>
<td>0.392*</td>
<td>0.027</td>
</tr>
<tr>
<td>Internal</td>
<td>2.10 (0.94)</td>
<td>1.88 (1.05)</td>
<td>0.168</td>
<td>0.258</td>
</tr>
<tr>
<td>Hot issues</td>
<td>1.42 (0.70)</td>
<td>1.28 (0.62)</td>
<td>0.175</td>
<td>0.339</td>
</tr>
</tbody>
</table>

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

Conclusion

This study designed a summarization game by incorporating keyword selection and auction mechanism. Although summarization is essentially a personal learning activity, group-based games may expend students’ horizon of reading. Furthermore, the auction mechanism provides students with opportunities to compare their evaluation of keywords with others, further stimulating students’ way of thinking. In this vein, auction games, which are rarely adopted in education, provide students with opportunities to express and quantify their ideas. By doing so, they may examine their own understanding further. In a sense, the game limits students’ selection of keywords, which allows them to make decision by evaluating the importance of keywords. However, even if students do not acquire important keywords, they still could replace them with synonyms or related words. More importantly, they may add more important ideas to compose summaries without addressing redundant or unnecessary information.

Most educational games focus on well-structured subject domains, probably because the knowledge in these subject domains is quantifiable. However, the knowledge in ill-structured subjects, like language, cannot be easily gamified, except basic facts. Instead, some researchers attempted to create a simulated, augmented, or authentic environment so that students can experience the use of languages in situated environments. The game rules of text games are relatively hard to design because games should fairly deal with human creations (e.g., text, speech, motions). This study provides a feasible example to start up such creation games. Furthermore, this game adopts a word segmentation technique to extract keywords. By weighting these keywords, one’s evaluation becomes quantifiable and human creations can be fairly gamified.

Besides, this study not only demonstrates students’ improvement of summary writing, but also explores the reading problems of graduate students. Specifically, they may lack the abilities of structure analysis, idea integration and argumentation. Unlike children, adult readers usually have specific intentions and prior knowledge. Even for first year graduate students, after a while, they likely have advanced academic needs, which may be supported by advanced technology. For example, after summarization, automatic essay scoring could be used to provide students with personal feedback for improving their writing. Alternative, their summaries could be compared with other players’, so that they could do self-reflection. On the other hand, text mining could be applied to automatically retrieve expert knowledge from research papers given specific research issues. By doing so, learning activities for graduate education like iAbstract could be extendable to other issues, other subject domains or even other ages.
Limitations and future works

Because this was an exploratory study, there were several limitations in this paper. First, our study adopted the summarization approach of McNeil and Donant (1982) as the design framework. The iAbstract game is only an implemented example of the framework. There are still many feasible ways to gamifying the process of summarization, while many further issues will emerge in these new ways. Furthermore, the auction mechanism provides students with interactions among players to facilitate their evaluation of the information in articles and the composition of short words for creating topic sentences. Such interactions, which may invoke or hinder students’ thinking, needs further investigation. Second, the game designed in this study initially incorporated some techniques of natural language processing to support the gameplay. Recently, the technologies of online text analysis become more mature. For example, Latent Dirichlet Allocation can be applied to dynamically analyze students’ topics and emotions in games from students-generated texts. Third, this study lacked rigorous experimental design, which required at least a control group and longer time. For this reason, the results may not be overgeneralized to other contexts. The learning effects of this educational game are also not included in this study and need to be examined in the future.

Acknowledgements

This study was funded by the National Key Research and Development Program of China (No. 2018YFB1004504), Hubei Provincial Science and Technology Program of China (No. 2017AKA191), and Self-determined Research Funds of Central China Normal University from the Colleges Basic Research and Operation of Ministry of Education of China (No. CCNU16GD003).

References


Pedagogical Change in Mathematics Learning: Harnessing the Power of Digital Game-Based Learning

Siew Pei Hwa
Department of Multimedia Design and Animation, Faculty of Creative Industries, Universiti Tunku Abdul Rahman, Malaysia // siewph@utar.edu.my

ABSTRACT
Technological applications, especially the use of multimedia courseware have become more common in today’s education, stimulating innovative approaches in teaching and learning. With the growing interest of integrating information and communication technologies (ICTs) into learning and teaching, the use of multimedia technology and digital games offered an alternate method of instruction. Multimedia objects play an important role in the classroom setting because of its ability to provide a virtual environment for learners to effectively acquire knowledge. This paper presents part of the findings of a research project that focused on the advantages of multimedia technology and the benefits of digital game-based learning. By using sample lessons from an interactive multimedia courseware called “DigiGEMs,” this paper emphasizes the use of digital games as a vital tool in mathematics learning. The study sets out to examine if a positive attitude exists among young learners towards the learning of mathematical concepts. The DigiGEMs is targeted at primary school children aged between 7 to 9 years old, who can practice mathematical thinking skills in an appealing manner. This paper also describes the efficacy of using multimedia and game-based approaches to motivate mathematical learnings among Primary 1 to 3 students. The research findings corroborate with the hypothesis that digital game-based learning is more effective than traditional class-based learning in acquiring mathematical knowledge. It is hoped that the discussion in this paper will encourage other researchers to not only conduct further studies on children’s learnings in the context of digital game-based learning environment in mathematics, but also to adopt game based-learning for other subjects such as science and languages in which the above elements (i.e., multimedia objects and digital games) are cohesively and dynamically integrated to optimise the teaching and learning processes.

Keywords
Mathematics learning, Digital game-based learning, Edutainment, Computer-based learning, Multimedia

Introduction
The advancement of information and communication technologies (ICTs) has triggered significant changes in all levels of education, from kindergarten right up to tertiary level. Education has evolved from the use of traditional methods of instruction to the use of ICTs for instruction. Kop et al. (2004, as cited in Shafie & Mansor, 2009) asserted that teaching in the digital age is no longer telling, and learning is no longer listening. Shafie and Mansor (2009) added that “the ideal learning environments for digital learners are rich learning environments that enable and support learners to learn independently and collaboratively” (p. 70). Malaysia has the second highest growth rates for e-learning products in the world, at the record of 39.4%, which is more than four times the worldwide aggregate growth rate (Sawahel, 2013). Clark and Mayer (2011) noted that e-learning involves the use of instructional methods (such as examples and practice) and multimedia elements (such as pictures and videos) to assist learning by delivering content which is pertinent to the learning objective.

Ever since, the introduction of e-learning has gradually become an important facilitator in the instructional process. Children of today are different from children of yesteryears. According to Kivunja (2014), Marc Prensky has referred 21st century children born after 1980 as “Digital Natives.” Kivunja (2014) added that Marc Prensky described digital natives as children who have grown up in a world surrounded by computers, videogames, digital music players, video cams, cell phones, and many other modern technological toys and tools. Prensky (2001) stated “there is no reason that a generation that can memorize over 100 Pokemon characters with all their characteristics, history and evolution can’t learn the names, populations, capitals and relationship of all the 101 nations in the world.” (p. 5). This statement clearly shows that children can actually learn and memorize things, even up to a list of 100 items, provided they are interested. According to Kirikkaya, Iseri and Vurkaya (2010, as cited in Liu & Chen, 2013), many researchers have scrutinized the effectiveness of digital technology in the promotion of learning, and many previous studies have demonstrated that learning motivation and efficiency can be enhanced through educational games. Omrod (2014) claimed that motivation often leads to improved performance. Past studies (e.g., Gottfried, 1990; Schiefele, Krapp, & Winteler, 1992; Walberg & Uguroglu, 1980, as cited in Omrod, 2014) had shown that students who are most motivated to learn and excel in classroom activities tend to be high achievers. Omrod (2014) also added motivated learners often make a concerted effort to truly understand classroom material, to learn it meaningfully, and consider how they might use the educational materials in their own lives. Many previous studies (Kirikkaya, Iseri & Vurkaya, 2010,
as cited in Liu & Chen, 2013) also found that game-based learning is often more effective than traditional teaching methods in boosting learning motivation, inculcating active participation, and concentration among students, as well as improving their skills in understanding and solving problems.

Pant (2013) stated that with the recent developments in technology, educational entertainment (which is referred as education plus entertainment) is any entertainment content-designed to educate as well as entertain, with a high degree of both educational and entertainment value is known as edutainment. Zentai and Dombovari (2009) claimed that edutainment is not simply a mixture of education and entertainment, but a special kind of toys or computer games. It is difficult to define its characteristics but if the creators want to be successful, the entertainment part has to be more distinctive than the education part. There are differences between games for education and games for entertainment. Developers of educational game must target the desired learning outcomes, and then design a game to achieve that target. Educational games built on the science of learning (Federation of American Scientists, 2006). Pant (2013) revealed that edutainment seems to be the new pedagogical model promoting active social learning. This new form of pedagogy has unlimited potential to create motivated and active learners, rather than bore passive learners in the usual classrooms. Edutainment could be used to promote active and engaged learning. It can be a means of transforming a new generation of smarter learners. The rise of gamification in learning is an important trend for the future in which the epistemic value of games is one of the important areas of educational exploration. The advancement and popularity of computers and multimedia technologies have encouraged researchers to develop digital content and systems for mathematics courses. The present paper highlights the creation of authentic learning scenarios that engage students in solving real-world problems with mathematical knowledge, whilst adopting the edutainment concept. It also investigates the effectiveness of digital technology and games in the promotion of mathematics learning among Malaysian Primary 1 to 3 students.

The remainder of this paper is structured in five sections. The second section outlines the need for pedagogical change, followed by the third section that presents a discussion on the concept of digital game-based learning. The fourth section describes the creation of game-based learning environment for mathematics learning, while the fifth section presents the efficacy of digital game-based learning for mathematics learning. The last section concludes the paper.

The need for pedagogical change

Mathematics is perceived by societies worldwide as the foundation for scientific and technological knowledge, as well as for political and socio-economic developments (Githua & Mwangi, 2003, as cited in Mutai, 2016). Thin (2015) had also affirmed that mathematical knowledge is important to the workforce in the information and technology age. It also serves as a vital tool in many fields, including natural science, engineering, medicine, finance and the social sciences throughout the world. Okebukola (1992, as cited in Mutai, 2016) referred to mathematics as “the central intellectual discipline of the technological societies” (p. 1). With reference to review of past studies, Mutai (2016) concluded that mathematics is the basis of all sciences and technology and of all human endeavours” (p. 1). Aminu (1990, as cited Mutai, 2016) stated that mathematics cuts across all areas of human knowledge. Mutai (2016) postulated that before an individual can function well in the society, one must possess relatively good knowledge of mathematics especially in this era of technological age. Hence, strong mathematical skills are critical for many careers in modern technological society (Stankous & Buibas, 2015). Berlinksi and Busso (2013) had also asserted that mathematical competence is a fundamental skill for personal fulfilment, active citizenship, social inclusion and employability in the modern world. A lack of sufficient mathematical skill and understanding affects one’s ability to make critically important educational, life, and career decisions (Sherman, Richardson & Yard, 2015).

Githua and Mwangi (2003, as stated in Thin, 2015) noted that life without mathematics is an almost impossibility. Thin postulated that mathematics is one of the most important subjects from which people can learn several vital skills. These skills include the ability to identify and analyse patterns, logic and critical thinking, ability to see relationships and problem-solving skills. It also allows people to perform complex calculations using a set formula or mathematical procedure. Aminu (1990, as cited in Mutai, 2016) also claimed that mathematics is an essential nutrient for thought, logical reasoning and progress. Fromboluti and Rinck (1999) noted that mathematics helps children make sense of the world around them and find meaning in the physical world. Through mathematics, children learn to reason, connect ideas, and think logically. It is also about connections and seeing relationships in everything they do. According to Fromboluti and Rinck (1999), The National Council of Teachers of Mathematics (NCTM, the world’s largest organisation concerned with
mathematics education) has developed a set of mathematics concepts, or standards, that are important for teaching and learning mathematics. There are two categories of standards:

- **The thinking math standards**: It focuses on the nature of mathematical reasoning, which include problem solving, communication, reasoning, and connections; and
- **The content math standards**: This refers to specific mathematics topics such as estimation, number sense, geometry and spatial sense, measurement, statistics and probability, fractions and decimals, and patterns and relationships.

Since mathematical knowledge is important to all human endeavours especially in scientific and technological development, it is a core subject in schools all over the world. In Malaysia, Mathematics is a core subject that must be taken in primary school examination namely UPSR (Ujian Pencapaian Sekolah Rendah or Primary Schooling Achievement Tests) (Mahamad, Ibrahim & Mohd Taib, 2010). However, the Malaysian Examination Board (as cited in Mahamad, Ibrahim & Mohd Taib, 2010) had reported that more than 35 percent of primary school students failed the subject in the 2008 UPSR examination. There is an increase in the number of students who had a poor performance in Mathematics in UPSR examination. Of the 440,496 primary school students who sat for the 2016 UPSR examination, 27.0 percent of them attained grade D in Mathematics and 21.6 percent attained grade E (Bicara Rakyat, 2016). According to Sherman, Richardson and Yard (2015), students fall below their expected level of mathematics achievement for a variety of reasons. Abdul Gafoor and Sarabi (2016) noted that curricular materials, gap between learner and subject matter, memory ability, attention span and understanding the language of mathematics are major reasons for students falling below their expected level of mathematics achievement. Sherman, Richardson and Yard (2015) reported that when students were asked why they were not successful in learning mathematics, many people replied that they “never understood math,” or “never liked it because it was too abstract and did not relate to them difficult to relate.” (p. 5).

The literature indicates that many factors contributed to the poor performance in mathematics and math anxiety. Numerous studies (e.g., Cavanaugh, 2004; Greenwood, 1984, as cited in Stankous & Buibas, 2015; Makinde, 2014) related these issues to the instructional methods used by teachers as a significant influence to the students’ attitudes towards math. Stankous and Buibas (2015) added that the teacher presents math problems in a didactic manner in most classrooms, causing students to be disengaged. Sherman, Richardson and Yard (2015) also revealed that most of the time, students who are taught in a way that relies too heavily on rote memorisation isolated from meaning have difficulty recognizing and retaining math concepts and generalizations. Boaler (2008, as cited in Stankous & Buibas, 2015) stated that “the design and presentation of mathematics in schools foster math anxiety. The standard method of instruction in many math classrooms is a significant factor leading to poor performance” (p. 204). Wu and Lin (2016) had also indicated that children in Taiwan have not been interested in or even afraid of learning mathematics, which incidentally triggered them to put more effort in resolving them.

Rajoo (2013) affirmed that a firm grasp of basic mathematical knowledge acquired at lower levels is vital for a student to progress to upper classes in secondary schools. Wu and Lin (2014) stated that since mathematical learning in early childhood plays a significant role in children’s future development of mathematical concepts, logical thinking, and learning attitude; how to enrich young children’s mathematical learning experiences at the elementary stages is important. Mahamad, Ibrahim and Mohd Taib (2010) noted that teaching and learning mathematics have significant challenges worldwide in terms of delivery methods and student’s participation. Bandura (1977, as cited in Saleh & Abdul Rahman, 2016) asserted that each student is unique in their acquisition of knowledge. Students’ poor performance may not necessarily be due to their non-possession of skills, but more to the fact that they have not been able to optimize their skills. Traditional and formal education may not have been able to engage students with different styles of learning. In any learning situation, students’ feelings, attitudes and motivation to learn are important to determine their learning performance or knowledge acquisition. According to Felder and Brent (2005), “students have different levels of motivation, different attitudes about teaching and learning, and different responses to specific classroom environments and instructional practices. The more thoroughly instructors understand the differences, the better chance they have of meeting the diverse learning needs of all of their students” (p. 57). Traditional classroom teachers have been concerned about students’ disengagement and lukewarm interest in studies. Damian and Duguid (2004, as cited in Hung, Huang & Hwang, 2014) reported that the application of multimedia and enjoyable tasks in acquiring mathematical concepts could assist students in mathematical learning and help them relate the concepts to their daily lives.
Computer-based learning (CBL) and its potential in teaching and learning environments

The difficulties in the learning of mathematics are well-known. Students often perceive mathematics as a difficult subject (Ngirande & Mutodi, 2014; Stodolsky, Salk, & Glaessner, 1991, as cited in Ku et al., 2014). It often evokes feelings of stress, anxiety and fear (Atallah, Bryant & Dada, 2009, as cited in Ngirande & Mutodi, 2014). Makinde (2014) revealed that mathematics has been proved to be a fearful subject probably due to its demand for higher-order thinking skill. Makinde (2014) also claimed that poor presentation skill coupled with the unfriendly attitude of some teachers has worsened the situation; resulting in many students losing interest in mathematics. Castro et al. (2015) noted that recent studies emphasized the relationship between motivation and learning mathematics. Pantziara and Philippou (2014, as cited in Castro et al., 2015) pointed out that “enhancing students’ motivation in the mathematics classroom is an important issue for teachers and researchers, due to its relation to students’ behaviour and achievement” (p. 236). Makinde (2014) feels that the effectiveness of teaching depends on a variety of factors: the use of relevant teaching aids, appropriate teaching methods, and the use of a student-centred approach in the delivery of teaching procedure. Fadel and Lemke (2008) noted that educators are in constant search for more efficient and effective ways to advance student learning.

Past studies (e.g., Copley, 2010; Diamond & Hopson, 1998, as cited in Wu and Lin, 2016) indicated that new skills acquired in a fun and effortless environment in the early learning stages can positively stimulate and enhance young children’s potential in exploring the world. Through the use of engaging and adaptable material, computer-based learning (CBL) has the potential to provide a highly motivating learning experience. However, certain features must be presented in CBL to ensure its effectiveness in offering support to delivery and learning in the classroom. The main features of CBL include (Clarke, 2001):

- presentation and combination of multiple media (graphics, still images, sound, animation, video, colour and text);
- adaptability (able to change to suit the learners’ needs, style and pace);
- dynamic display (windows, scrolling and hypertext links);
- memory (record learners’ reactions, test results and patterns of learning);
- patience (computers make no judgment if learners need many attempts to understand the content);
- tirelessness (never need to take a break or go on holiday); and
- interactivity (able to respond to learners’ reactions, behaviour and choices).

Clarke also highlighted that the critical feature of CBL which makes it different from other media is interactivity, which is the power of the computer to engage, communicate with and adapt to the learner. Learning is an active process, so simply presenting information on the screen is unlikely to be successful. Learners must be able to interact with the content by making choices and receiving feedback. CBL serves to establish more effective learning situations than traditional teaching and learning. CBL in Malaysia has emerged as an instructional technology tool with the potential to surpass the conventional methods, which are, for example, usually carried out by chalk and talk (Nordin & Fatimah, 2011, as cited in Ong et al., 2013). Ong et al. (2013) pointed out that many Malaysian educators have harnessed the power of computer technology in helping students to learn Mathematics; for instances, educational software such as interactive multimedia application and freeware are available for instructors to use in teaching and learning.

Interactive multimedia application provides a mechanism to revolutionise mathematics instruction. The use of new technologies can facilitate instruction in the classroom – an increasing trend in computer use at the primary school level has resulted in positive changes in the learning process. Multimedia programmes are designed to support and accelerate the learning process. Educators have been using multimedia as a teaching tool for several years. This is because multimedia is able to offer the experience of listening, looking and doing in a computer-mediated setting. It can be exciting, interesting, motivating, and help students achieve understanding in new ways. The use of sound, photographs and video enables the user to experience authentic situations which may not be possible with conventional methods of instruction (Oliver, 2000). This means that computer and multimedia technologies have simply provided the vehicle for teaching and learning, with the potential to overcome the limitations of traditional media in supporting the prospect to provide learning environments with strong visual elements (Nordilah, 2010, as cited in Ong et al., 2013). The Malaysian government believes that using courseware which has the positive edge of providing students with a greater level of individualised instruction can increase student engagement and motivation (Barrow, Debraggio & Rouse, 2008; Chong, Horani & Daniel, 2005, as cited in Ong et al., 2013).

Numerous past studies (e.g., Arif & Lim, 2009; Noordin & Fatimah, 2011; Lee, 2010; Mahmud, Teoh et al., 2009; Zuraini & Fatimah, 2010) had proved the positive effect of CBL in learning mathematics (Ong et al., 2013). The result of a study done at the University of Texas by Metcalf (1997, as cited in Makinde, 2014; Sutton,
2013) confirmed that people remember 10 percent of what they read; 20 percent of what they hear; 30 percent of what they see; 50 percent of what they see and hear; 70 percent of what they say; and 90 percent of what they do and say. Makinde maintained that play-way method could be a way of achieving effective teaching of mathematics as it involves the act of seeing, hearing, saying and doing, which makes it easier to remember. Past studies (e.g., Ke & Grabowski, 2007, as cited in Ku et al., 2014) found that digital games had potentials to enhance students’ learning performance. Rieber (2005 as cited in Abdul Rabu & Talib, 2017) defined digital game as “a game that involves interaction between humans and multimedia elements, but requires interaction at a higher level than interactive multimedia applications.” (p.62), whereas Mat Sin, Talib, and Norishah (2013 as cited in Abdul Rabu & Talib, 2017) described digital games in the context of education as “an interaction of one or more players with multimedia components to educate new knowledge and skills to the players or students.” (p. 62). Kuo (2007, as cited in Hung, Huang & Hwang, 2014) reported that digital games could provide a learning environment that attracts students’ interest in learning mathematics. Therefore, embedding digital games into mathematics learning may be a possible solution to enhance students’ learning performance, and practice mathematical thinking skills in an appealing manner.

Learning and teaching through play-way method in the learning environments for children

Children love to play and it is their natural instinct (Yohannan, 2015). Pant (2013) noted that during the early years of childhood, the play-way method for learning is universally used. The concept of play-way was first introduced by Friedrich Froebel (Argawal, 2016). Preerzada (2016) noted that Friedrich Froebel has made a valuable contribution to educational theory and practice. Froebel was the first educator to establish that self-activity is a core component of teaching-learning in the classroom. Mareshwari (2013) has highlighted several principles of play-way method. Some of them are as below:

- It is based on activity based learning that stimulates creative skill and self-expression;
- It is life oriented and sets school as a second home for the children;
- The needs of the children can be fulfilled and increased through play-way method;
- It narrows down the gap between the children and the teachers;
- It helps to prolong the memory skill; and
- It provides opportunities to all children to participate

According to Robinson (2012) and Zav (2017), the play-way method which encompasses the methods of teaching such as kindergarten method, Montessori Method, Dalton plan, Heuristic Method of Armstrong, Project Method, and so forth are all attempts to imbibe play spirit in education. Preerzada (2016) further explained that Froebel’s principle of play-way has been accepted by educators and has been adopted in teaching and learning processes at all levels of education from primary to university and in all subjects. Also, Makinde (2014) believed that play-way method of teaching is useful or applicable at the primary or kindergarten level of education, just as Fredrich Froebel advocated “guided play as the best way for a child to learn”. Robinson (2012) and Zav (2017) asserted that play-way in education insists on child-centred education, rather than on subject matter. It introduces the spirit of play in all educational instruction. The play-way method also emphasises the spirit of freedom, spontaneity and enjoyment introduced in all school work (Zav, 2017). Robinson also stressed that education should be fun and not forced. Play-way method advocates educating children through activities that engage, entertain and instil learning. Educational games serve a good purpose in this direction. Besides the ability of making learning more interesting, educators and researchers have been exploring the pedagogical potential of educational games (Jong et al., 2008).

The concept of digital game-based learning

The teacher-dominated and knowledge-centred mathematics teaching is rather traditional, especially from the western perspective (Biggs, 1998; Zheng, 2006, as cited in Yang, 2013). Teaching strategies underpin all learning in the classroom. They determine what is learned and the nature of the interactions between students and teachers (Berlinski & Busso, 2013). The main problem pertaining to the traditional way of conducting mathematical learning is a lack of motivation. Children do not seem to be motivated in learning the subject because they do not see the use of mathematics concepts in everyday life. Prensky (2003) highlighted that “A sine qua non of successful learning is motivation.” (p. 1). Prensky (2003) stressed that much of the content that needs to be learned by students these days does not directly motivate them, irrespective of whether the learners are in schools, colleges, corporations, professions, or even the military. The word “boring” or “dry” often crosses their lips. Demotivation may cause the real problems in getting highly-stimulated students to learn. Therefore, many people have tried to merge the content of learning and the motivation of games. Digital games have been
proposed as one of the effective tools if they were combined with education to promote student engagement and joyful learning process (Prensky, 2003). Past studies (e.g., Ke, 2008; Papastergiou, 2009; Prensky, 2002 as cited in Che Pee, 2011) pointed out that motivation is one of the most important features in learning and the element of “fun” in a game situation will be a key factor in motivating current and future learners. In the review of literature, many researchers (e.g., Bixler, 2006; Klien & Freitag, 1996; Shaffer, 2006; Thatcher, 1990, as cited in Che Pee, 2011) had made an assumption that the rationale of using games for learning is that games are intrinsically motivating. Zikas et al. (2013) also asserted that the principal aim of applying games in education is to increase students’ engagement and motivation.

Shearer (2011) defined digital game-based learning as “… is an instructional method that incorporates educational content or learning principles into video games with the goal of engaging learners.” (p. 6). Shearer further explained that digital game-based learning draws upon the constructivist theory of education. Digital game-based learning involves a fusion/involve the integration of educational content and computer or video games and can be used in almost all subjects and skill levels. In concurring with this, Prensky (2001, as cited in Al-Washimi, Blanchfield & Hopkins, 2015) also postulated that the incorporation/ injection of computer games would offer a more effective approach than the conventional one:

- Being “digital natives,” millennium students need a different learning approach which can engage their inquisitive minds; and
- The conventional teaching and learning approaches used in the classrooms do not work and need to be replaced.

Eurydicev (2011 as cited in Berlinksi & Busso, 2013) stated that “Moving away from the traditional teacher-dominated way of learning, active learning approaches encourage pupils to participate in their own learning through discussions, project work, practical exercises and other ways to help them reflect upon and explain their mathematics learning.” (p. 2). Gros (2007) asserted that digital games are user-centred; they can promote challenges, cooperation, engagements, and the development of problem-solving strategies. The design of a learning environment built on the educational properties of games can be an appropriate way to improve learning. Wu and Lin (2016) indicated that “a child-centred and child development-based mathematics curriculum could not only help teachers see how the kids learn and understand mathematical concepts, while attaining these concepts, but also promote their high-level mathematical development” (p. 844). Past studies (e.g., Shaffer, Squire, Halverson & Gee 2005; Klawe, 1999; Williamson, 2009) proved that computer games can be used to support children’s learning (Che Pee, 2011). McAnallen (2010, as cited in Stankous & Buibas, 2015, p. 208) noted: “Today’s teachers can and must learn to serve as mathematical linguists, artists, musicians, and dancers so they can instil the joy and wonder of mathematics to generations of students in the future” (p. 208).

According to Prensky (2003), digital game-based learning is an integration of educational content and computer games. Sayed Yusoff, Tan, and Muhammad Zaffwan (2013 as cited in Abdul Rabu & Talib, 2017) in their systematic research review reported that new digital game innovations have begun to be applied and their effectiveness has been assessed on mathematics, arithmetic, chemistry and early childhood education. Prensky (2001) asserted that games have the power of engaging people. They are fun and provide interaction, interactivity, problem solving, story and other elements that give the user involvement, structure, motivation and creativity, among other benefits. There were extensive empirical evidences which show the positive effect of educational computer games on students’ performance in mathematics, science or the military (McFarlane et al., 2002). Zentai and Dombovari (2009) pointed out that interactivity is one of the most important characteristics of these games; it is crucial that the practical element of interactivity is embedded in the rules of the game.

In recent years, digital games have assumed an important place in the lives of children and adolescents. The power of gaming is particularly relevant to today’s students who have grown up with interactive digital technologies. These digital natives are not merely technologically savvy but approach lives differently as they have seamless integration of digital technologies (e.g., computers, the Internet, smartphones, mobile messaging) in their daily activities. Digital game-based learning is emerging as the new dynamic model of e-learning (Squire, 2005) due to various reasons. The chief reason is that people are usually bored with traditional learning whereas games are motivating and engaging. Many people believe that educational games are effective because they motivate children to actively engage in a learning as they play. Engagement and motivation are interesting benefits of the use of games in learning (Gros, 2007).

The use of educational games within lesson plans is an increasingly common practice for teachers and educational technologists. There are several pedagogical approaches that can be used in and/or around a game, such as learning by doing, learning from mistakes, goal oriented learning, role playing and constructivist learning (Prensky, 2001). Integrating computer assisted and game-based learning into the mathematics learning
process can greatly enhance the learning experience and significantly increase success rate in knowledge acquisition for a majority of students. Games introduce/teach specific content and skills in a user-friendly environment where students are able to play, try, make mistakes and learn. At the same time, game-based e-learning approach may foster a positive attitude among young learners towards the mathematical sciences. Snow (2016) claimed that digital games encourage players to take on active roles, solve problems, and develop and draw upon skills that incorporate active learning. This type of active learning is frequently not found in the traditional classroom setting. Most of the past studies proved that games increase student engagement during learning activities (e.g., Annetta et al., 2009; Fowler & Cusack, 2011; Kiili et al., 2014; Rai & Beck, 2012a; Rai & Beck, 2012b, as cited in Marlund, 2015; Chen, 2017; Huizenga, 2017; Schaaf, 2012), motivation and learning achievements/outcomes (e.g., Chen, 2017; Huizenga, 2017; Snow, 2016) or that they provide more efficient teaching in the styles of newer pedagogical paradigms (e.g., Gee, 2003; Mayo, 2007; Shapley et al., 2011, as cited in Marlund, 2015). Past studies also indicated that game-based learning has positive effects on mathematics learning (Drigas & Pappas, 2015; Ferguson, 2014; Katmada, 2014; Lin et al., 2013; McLaren et al., 2017).

**Creating a digital game-based learning environment for mathematics learning**

The notion that children learn by constructing their own knowledge is highly popular among educational theorists. Children ought to be active, not passive, in the learning process. They ought to be doing something, not merely watching it. Multimedia technologies offer children the opportunities of learning “actively” by allowing them to construct knowledge in the form of interactive multimedia documents (e.g., multimedia stories). In this study, a game-based learning approach is proposed by integrating educational mini games into the mathematics content to optimise the teaching learning process among Primary 1 to 3 students.

There has been a paradigm shift in education from a didactic model of instruction to a constructivist model that emphasises more on the role of an active learner (Jong et al., 2008). This paper discusses a pedagogical intervention designed to give primary school students a more active role in the classroom. It is suggested that using multimedia learning objects and computer aided learning can vastly enhance the effectiveness and positive impact of the mathematics learning process among young learners. This study seeks to design a digital game-based learning environment using an entertaining and interesting courseware called DigiGEMs (Digital Games for Education in Mathematics) that aims to teach mathematics among Year 1 to Year 3 Malaysian primary school students. The learning environment caters for different learning styles among young learners aged between 7 to 9 years old. The DigiGEMs reflects a new, multimedia mediated game-based approach to foster a positive attitude towards mathematical science in children. It is an interactive educational tool that centres on the children, who will construct their mathematical knowledge in a gradual fashion. Based on the mathematics syllabus of the Malaysian Ministry of Education for Year 1 to Year 3 primary level, there were 16 lessons built into the DigiGEMs:

- Lesson 1: Quantity Comparison,
- Lesson 2: Numbers Within 0 To 20,
- Lesson 3: Ordinal Number Within 20,
- Lesson 4: Numerals and Words Within 20,
- Lesson 5: Arranging Numbers Within 0 To 20,
- Lesson 6: Addition Within 20,
- Lesson 7: Subtraction Within 20,
- Lesson 8: Numbers Within 100,
- Lesson 9: Counting Numbers Within 100,
- Lesson 10: Arranging Numbers Within 100,
- Lesson 11: Addition Within 100,
- Lesson 12: Subtraction Within 100,
- Lesson 13: Solving Daily Problems,
- Lesson 14: Shape and Space,
- Lesson 15: Introduction To Time, and
- Lesson 16: Naming Days and Months

Each lesson encompasses the introduction of a topic and detailed explanation of each topic using a variety of multimedia objects such as graphics, images, animation, audio and video (see Figure 1), as well as three game-based drill activities (see Figures 2 through 9).
The DigiGEMs encompasses a wide range of enjoyable computer-based activities. It specifically refers to educational software that incorporates some popular electronic game elements such as a high degree of interactivity, exploration, puzzles, challenges, scoring, graphics, music, sound effects, or narratives. The subsequent paragraphs discuss some sample screenshots from 5 of the 16 lessons (Lessons 1 to 4 and 15) in DigiGEMs that adopts educational games approach in enhancing children’s understanding of mathematical concepts.

**Example 1 - Lesson 1: Quantity comparison**

The game built into Lesson 1 is a “Card Matching” mini game to test the understanding of “Quantity” concepts of the children. As shown in Figure 2, 16 cards appear with different number of fruits; two of the cards have the same number. The outline of each card is in black colour before the game starts.

**Figure 1. Sample screenshots from Lesson 1**

**Figure 2. Sample screenshot from the “Card Matching” mini game in Lesson 1**

**Figure 3. Sample screenshot from the “Card Matching” mini game in Lesson 1 showing the yellow coloured outline of the card**
When the game begins, learners need to find and match the two cards with same quantity. The rules of playing this “Card Matching” game are as follows:

- When learners click on one of the cards, the outline of the card will be highlighted in yellow colour as shown in Figure 3.
- Then, the learner needs to find another card with the same quantity (in this case, the number of fruits is 6) to match with it. If learners match the cards correctly, both cards will disappear (see Figure 4). If the first selected is incorrectly matched (i.e., different quantity), the yellow-coloured outline of the card will revert to black outline as shown in Figure 2; both cards remain on the screen.

![Figure 4. Sample screenshot from the “Card Matching” mini game in Lesson 1 showing the disappearance of two cards with same quantity of fruits (in this case is 6)](image)

**Example 2- Lesson 2: Numbers within 0 to 20**

The game embedded in the exercise module of lesson 2 is “Balloon Shooting” activity. After learner accesses to the gaming page as shown in Figure 5, a number in written word will appear on the cloud (e.g., Ten) and some balloons with different numbers (e.g., 10, 17, 1, 11, 18) will float in the air. Learner needs to move “shooting” cursor to shoot the correct answer based on the number in written word appear on the cloud (e.g., Ten = 10). If the answer is correct, the correct balloon will burst.

![Figure 5. Screenshot from “Balloon Shooting” game in Lesson 2](image)

**Example 3- Lesson 3: Ordinal number within 20**

In this drill activity, the learner needs to drag and drop the number in written words (e.g., Thirteenth) into the correct answer boxes as shown in Figure 6. When the learner presses the generator button, a new number in written word will be generated. The learner needs to drag the word and drop into the appropriate answer box to match it with its correct answer (e.g., Thirteenth = 13th). After the learner drops an answer into a box, a mascot will appear and validate the learner’s answer as shown in Figure 7.
Example 4- Lesson 4: Numerals and words within 20

The “Join the Dots” mini game is one of the games embedded in Lesson 4. In this activity, the learner needs to join all the dots by following the sequence of numbers to reveal a picture. If a learner successfully joins all dots in the sequence as shown in Figure 8 (Left), a screen as depicted in Figure 8 (Right) is displayed. The aim of this activity is to test the children on the sequence of numbers within 20.

Example 5- Lesson 15: Introduction to time

The “Time Matching” mini game is one of the educational games integrated into Lesson 15. When the game starts, the learner needs to match the digital clock (e.g., 5:00) with the same time shown on the analog clock, as depicted in Figure 9 (Left). The learner matches by dragging and dropping the digital clock into an appropriate box below the analog clock. After all the digital clocks have been matched with the appropriate analog clocks, the learner needs to press the “OK” button to obtain a validation of the answers. A “Check” sign (✓) appears beside the correct answer, while a “Cross” sign (✗) will appear beside the wrong answer as shown in Figure 9 (Right). However, the learner will be given opportunities to re-match the analog clocks with wrong answers until he/ she accomplishes all the matching pairs of clocks.
Figure 9. Sample screenshots from the gaming page of “Time Matching” mini game in Lesson 15

Although there are ample multimedia courseware and online courses that adopt game-based format for learning mathematics at Primary school level that are available in the market, most of the instructional materials are developed abroad and do not comply with the local needs and educational values. The mathematical content with digital games format does not meet the requirements of the Malaysian mathematics syllabus from the Malaysian Ministry of Education (MoE). Thus, those online courses for learning mathematics using educational games are not suitable as a supplementary tool in the teaching and learning of mathematics for Primary 1 to 3 students in Malaysia. DigiGEMs is designed to meet the MoE’s requirements and will help the Malaysian primary school students learn mathematics effectively.

From the review of extant literature, it is found that digital game-based learning could be a good approach for improving students’ motivation and achievement in mathematics. Hwang et al. (2012, as cited in Hung, Huang & Hwang, 2014) depicted a digital game-based learning activity engages students in the process of problem solving or knowledge acquisition when faced with the challenges presented by the game. Hwang et al. (2012, as cited in Hung, Huang & Hwang, 2014) also assumed that by integrating digital games into instructional materials, students’ learning motivation would be enhanced because of the challenging and enjoyable nature of the games. The integration of learning with gaming makes mathematics learning more fun which in turn motivates students and helps them pay attention and stay focused on the subject (Science Education resource Center”, 2018). Tan et al. (2008) noted that digital games are relatively more fun, attractive and motivating to use compared to other modern forms of new media and therefore have greater potential in education. The DigiGEMs may not only stimulate children’s interest in learning through the power of digital games, but also assist them in understanding mathematics concepts expediently and provides a delightful multimedia experience.

The efficacy of digital game-based learning environment for mathematics learning

Hung, Huang and Hwang (2014) reported that many studies (e.g., Gee, 2003; Prensky, 2006; Squire & Jenkins, 2003) have proven the benefits of learning with educational games versus traditional approaches. In this study, the author attempts to examine the value of digital game-based learning environment by comparing the learning performance of students who were exposed to the digital game-based format verses those who were exposed to the traditional-based learning approach in the classroom.

Research participants

All the participants were Primary 3 school students. An experiment was designed to examine the effects of educational games with a focus on mathematics learning achievement. The participants were recruited via convenience sampling, i.e., 10 students were randomly selected from a class to participate in the educational games experimental group, while the other 10 students were regarded as the control group. The sample size adequacy should be determined before data collection is conducted. Scholars (e.g., Creswell, 2014; Fraenkel, Wallen, & Hyun, 2014) commonly suggest that carrying out experimental studies with at least 30 participants in one group is necessary, otherwise there will be inadequate power and the assumptions required of specific statistical tests used in the study will not be met. However, some scholars such as Bouma and Atkinson (1995, as cited in Shiratuddin, 2002) noted that if the population that is to be sampled is fairly homogeneous (i.e., the relevant characteristics are fairly and evenly distributed), a smaller sample can be relied upon than if the population is highly variable. In this study, the requirements of the sample were:

- experience in using the same printed mathematics textbooks in school since Primary 1; and
- basic knowledge of the 16 mathematics lessons covered in the DigiGEMs.
Since these requirements were easily met, it is possible to show that sample characteristics such as different gender, ethnicity, and learning experience are fairly homogeneous in any sample of students. Hence, a small number of participants is assumed to be adequate. To this end, the students in the experimental group studied using the digital game-based learning approach, while those in the control group used the traditional “chalk and talk” instruction approach to provide a comparison. It was implemented by giving well-structured lessons in the classroom that complemented the prescribed printed textbook and traditional printed workbook. During the empirical study, the control group continued to learn mathematics lessons through regular classroom instruction. Pre- and post-tests were undertaken to collect data and information on the impact of educational games in mathematical learning.

Research instrument and procedure

In this study, a pre-test and post-test-designed quasi-experiment was conducted. Two sets of achievement tests (i.e., pre-and post-test) were developed by the author which modified from several past teacher-made mathematical tests that were designed by experienced primary school mathematics teachers (i.e., teachers with teaching experience in mathematics for more than 5 years). Kinyua and Okunya (2014) noted that teacher-made tests are generally valid and reliable. Each test covered 15 multiple-choice questions on the selected lessons (i.e., Lessons 1 - 8) presented in DigiGEMs. Examples of questions are as follows:

1. **Choose the correct answers.**

   \[ \text{A. 17} \quad \text{B. 18} \quad \text{C. 19} \quad \text{D. 20} \]

   \[ \text{Answers: A. 17} \quad \text{B. 18} \quad \text{C. 19} \quad \text{D. 20} \]

2. **Choose the correct answers.**

   \[ \text{A. 76} \quad \text{B. 67} \quad \text{C. 89} \quad \text{D. 58} \]

   \[ \text{Answers: A. 76} \quad \text{B. 67} \quad \text{C. 89} \quad \text{D. 58} \]

3. **Choose the correct answers.**

   \[ \text{TWENTY} \]

   \[ \text{Answers: A. 76} \quad \text{B. 67} \quad \text{C. 89} \quad \text{D. 58} \]

A pre-test was conducted before the experiment to evaluate the prior mathematical knowledge of the students. At the beginning of the study, both experimental and control groups were asked to respond to the first set of achievement test to measure their prior achievement in selected lessons. Before the intervention started, the experimental group in the game-based learning environment was given a 20-min introductory session on how to use the DigiGEMs. The experimental group then was given a 30-min “warming up” session to familiarize them with computer and DigiGEMs. The nature of the study and the conditions of the class did not allow the experimental group to have learning sessions in the computer laboratory, therefore each student in the experimental group was given a piece of DigiGEM to practice during off-class hours with parents’ guidance. The experimental group worked on the DigiGEMs at their own pace for 5 weeks. It was assumed that the control group did not have a chance to practice the DigiGEMs. Once the experimental group completed working through the lessons, both the experimental and control groups were tested using the second set of achievement test. To measure their achievement, both the experimental and control groups were given an achievement test, that is, a post-test consisting questions identical to those in the pre-test assessment. Table 1 depicts the data collection procedure.

<table>
<thead>
<tr>
<th>Group</th>
<th>Pre-test</th>
<th>Treatment</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control (10 subjects)</td>
<td>Achievement test set 1</td>
<td>Traditional classroom instruction included mainly “chalk-and-talk” instruction supported by the prescribed printed textbook and traditional printed workbook</td>
<td>Achievement test set 2</td>
</tr>
<tr>
<td>Experimental (10 subjects)</td>
<td>Achievement test set 1</td>
<td>Blended learning environment with game-based approach through the use of DigiGEMs</td>
<td>Achievement test set 2</td>
</tr>
</tbody>
</table>

Marks obtained through the pre- and post-tests conducted were calculated to measure students’ achievement in mathematics knowledge acquisition before and after intervention (i.e., learning mathematics in a blended learning environment with game-based approach through the use of DigiGEMs). Pre- and post-tests marks were used instead of formal exam marks since it was impossible for the author to conduct a formal final exam before and after treatment to monitor students’ performance and achievement.
Data analysis and research findings

Before the intervention started, it was assumed that there is no significant difference between the experimental (adopted a blended learning environment with game-based approach through the use of DigiGEMs) and control (adopted traditional classroom instruction) groups at the time of pre-test as both adopted same teaching-learning approach (i.e., traditional face-to-face classroom instruction). The pre-test results are shown in Table 2.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean ± S.E</th>
<th>SD</th>
<th>t-value</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>10</td>
<td>8.90 ± 0.71</td>
<td>2.23</td>
<td>0.28</td>
<td>0.787</td>
</tr>
<tr>
<td>Control</td>
<td>10</td>
<td>8.60 ± 0.83</td>
<td>2.63</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Thereinafter, the analysis of data was done based on parametric approach using SPSS (Statistical Package for Social Science). Independent-samples t-test was used to test the following null hypothesis, $H_0$: There is no significant difference between game-based learning environment and traditional-based learning in the classroom

An independent-samples t-test was used to determine whether there is a statistically significant difference between the means in two unrelated samples (Usman, 2016). Elliott and Woodward (2007, as cited in de Winter, 2013) stated that “if one or more of the sample sizes are small and the data contain significant departures from normality, you should perform a non-parametric test in lieu of the t-test.” (p. 2). However, Usman (2016) claimed that non-parametric tests are usually less powerful than corresponding tests designed for use on data that come from a specific distribution. Therefore, even though this study involved a small sample size (i.e., $N = 10$), the parametric (independent $t$-test) was used since the sample characteristics such as different gender, ethnicity, and learning experience were fairly homogeneous in any sample of students. de Winter (2013) noted that some methodologists have cautioned against using the $t$-test when the sample size is extremely small ($N \leq 5$), however, de Winter asserted that some studies (e.g., Bridge & Sawilowsky, 1999; Janušonis, 2009) had suggested that applying the $t$-test on small samples is feasible.

The data collected was coded into SPSS programme for data analysis. The $H_0$ testing was to determine whether there was a significant difference between the experimental group and control group in Mathematical learnings. In relation to the testing of $H_0$, the independent-samples t-test was used to determine whether there is a significant difference between the mean scores of students’ achievements under two different conditions (i.e., digital game-based learning and traditional classroom instruction) among Primary 3 students. Through the post-test results as shown in Table 3, the $p$-value was 0.019 indicating that the data provides enough evidence to reject $H_0$ at a 0.05 significance level ($p < .05$). This explains that there was a significant difference between the achievements of the experimental and control groups on post-test. It implies that the digital game-based learning approach through DigiGEMs can affect the performance of students in mathematics learning. Therefore, there was strong evidence to support the hypothesis which stated that there was a significant difference between the means in two unrelated samples ($t \leq 5$), however, de Winter stated that “if one or more of the sample sizes are small and the data contain significant departures from normality, you should perform a non-parametric test in lieu of the $t$-test.” (p. 2).

The data collected was coded into SPSS programme for data analysis. The $H_0$ testing was to determine whether there was a significant difference between the experimental group and control group in Mathematical learnings.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean ± S.E</th>
<th>SD</th>
<th>t-value</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>10</td>
<td>11.90 ± 0.71</td>
<td>2.23</td>
<td>2.58</td>
<td>0.019*</td>
</tr>
<tr>
<td>Control</td>
<td>10</td>
<td>9.10 ± 0.82</td>
<td>2.63</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *$p < .05$.

The difference between the achievements of both the experimental and control groups in the post-test is clearly shown in Figure 10. Figure 10 reveals that none of the students in the control groups (traditional instruction) attained the value of “Very Good” compared to students in the experimental group (digital game-based learning or DGBL approach), which had 40 percent falling in the “Very Good” category. On the contrary, none of the students in the experimental group’s scores fell in the value “Poor,” while the control group had 20 percent. Meantime, the percentage of students in the DGBL approach that contributed to the value “Good” was slightly higher (50 percent) than the students using the traditional instruction approach (40 percent). In addition, 40 percent of students in the control group’s rating fell in the value “Average,” whereas the students in the experimental group was only 10 percent. Overall, the results indicate that using DGBL approach through DigiGEMs is more effective than traditional instruction approach in teaching and learning of mathematics at the elementary level.
The research findings from the current attempt are coherent with several empirical studies (e.g., Lee, 2010; Mahmud, Arif & Lim, 2009; Noordin & Fatimah, 2011; Teoh et al., 2009; Zuraini & Fatimah, 2010, as cited in Ong et al., 2013). These studies reported that CBL had an ability to improve the performance of students in mathematics. The research findings are also parallel with several studies which revealed positive effects brought about through digital games to support mathematics learning. The experimental results of Hung, Huang and Hwang (2014) showed that the game-based learning model had effectively promoted the students’ learning achievement, self-efficacy, and motivation in learning mathematics. The results of Ku et al. (2014) demonstrated that the digital games approach yielded better outcomes than the paper-based setting in both students’ confidence and performance. Ku et al. also reported that numerous past studies (e.g., Ke & Grabowski, 2007; Owston, Wideman, Ronda & Brown, 2009; Papastergiou, 2009; Kebritchi, Hirumi & Bai, 2010; Tsai, Yu & Hsiao; 2012) discovered that students in digital game-based learning settings had better learning achievement in mathematics compared to conventional instruction approach or face-to-face instruction.

Meantime, the results also correspond with those in previous studies which employed the use of digital game-based learning in various subjects. This proves that digital games have a positive impact on learning achievement. The research findings of Liu and Chen (2013) showed that the card game had significantly increased student’s scientific knowledge related to energy and means of transport, as well as enhanced learning motivation and learning effectiveness. Suh, Kim & Kim (2010, as cited in Ku et al., 2014) proved that students who used online role-playing games for English learning had better performance than those who received face-to-face instruction. Yien et al. (2011, as cited in Hung, Huang & Hwang, 2014) also reported the positive effects of computer games on students’ learning achievement in a nutrition course.

**Conclusion**

In view of the limitations of the traditional class-based learning approach, this study has explored the possibilities of introducing a game-based learning approach in mathematical learnings at elementary level. The aforementioned results suggest that the digital game-based approach seems to be a better than the conventional “chalk-and-talk” instruction. The design and development of the digital game-based learning environment in this study has benefited learners through the incorporation of the elements of multimedia mediated game-based learning that are conducive to mathematical learnings. Digital game-based learning is ideal for the purpose of adding a creative dimension to the lacklustre classroom environment. The DigiGEMs takes on the role of a very patient private tutor who is willing to repeat countless times per the learners’ needs. It also proved that mathematical learnings in a game-based learning environment can be achieved by utilising a variety of multimedia learning objects and mini educational games. Whilst educational games can provide a fun (Kebritchy, 2010, as cited in Al-Washimi, Blanchfield & Hopkins, 2015; Plensky, 2001), pleasurable, challenging, rewarding (Plensky, 2001) and engaging experience for students (Kebritchy, 2010, as cited in Al-Washimi, Blanchfield & Hopkins, 2015), game-based learning tools cannot replace classroom teachers in its provision for diverse students with different capabilities, personalities, and needs (Roe & Dickmeis, 2014, as cited in diverse Al-Washimi, Blanchfield & Hopkins, 2015). Therefore, the DigiGEMs can also be a supplementary tool in a blended learning environment that fuses class-based learning and e-learning activities, thus resulting in an effective instructional method that could optimise the mathematics learning experience.
Nonetheless, some improvements can be made to the present gaming model in learning environments in the future. Although the students can communicate with their peers, teachers and facilitators, the current educational games embedded as drill activities in the DigiGEMs are designed for individuals. Hwang et al. (2012) and Villalta et al. (2011) (as cited in Hung, Huang & Hwang, 2014) postulated that it would be more interesting and effective if the games were designed for students to learn mathematics collaboratively or competitively.

There are other factors that should be taken into consideration when designing the digital game-based framework to optimise the learning experience i.e., students’ different backgrounds, strengths, weaknesses, interests, ambitions, sense of responsibility, levels of motivation, and study approaches. In short, no two students are alike (Felder & Brent, 2005). Past studies indicated that students with different academic abilities performed differently (Colquitt, LePine & Noe, 2000, as cited in Ku et al., 2014). Al-Washimi, Blanchfield and Hopkins (2015) showed that children who engaged with the game achieve higher mathematics solving proficiency skills, whereas those who do not engage properly with the game performed worse in their post-tests. Thus, it is essential for researchers to put more emphasis on individual differences when designing new pedagogies, and analyse the effects of these pedagogies since students’ academic ability is an important variable that may affect students’ task performance (Ku et al., 2014). Further exploration to examine whether or not DigiGEMs could reinforce all students in mathematical learnings is much needed in the near future. Although the research outcomes attained in the current attempt indicate that the multimedia mediated digital game-based instructional materials have a positive effect on students’ achievement in mathematics, the study did not take into account the effect on students with different levels of academic ability; hence future studies could consider this.

In addition, this is a small-scale research where the findings could not be generalized to a larger population; the results in this research may not be sufficient to provide in-depth insights into Malaysian primary school students’ achievement in mathematical learnings. For future studies, the scope of the sample could be expanded to include more students and schools, and mathematics instructors would thus be able to make use of the results to better handle low-achievers and formulate remedial measures for them. It is hoped that the information and research findings obtained in current and future attempts will give rise to better strategies and measures for promoting student understanding and achievement in mathematics not only at elementary level in the Malaysian context, but also include students at higher education levels, different countries, different cultures, and so forth, thus contributing to a paradigm or shift change in mathematics education worldwide.

The author’s additional hope is that the discussion in this paper will inspire other researchers not only to conduct further studies on the inclusion of digital games to support mathematics learning, but also investigate how to create and examine the efficacy of a game based-learning context in other subjects such as science and languages in which the above elements (i.e., multimedia objects and digital games) are dynamically integrated to optimise the teaching and learning processes. More thoughts should also be given to treat game-based learning as a topic of serious research in the field of education.

Acknowledgements

The author would like to thank See Chee Wee for his work in developing the prototype of DigiGEMS. The author would also like to express her thanks to the students, parents and staff at a primary school for their involvement in the evaluation study and help in implementing the experiment.

References


Gamifying and Mobilising Social Enquiry-based Learning in Authentic Outdoor Environments

Morris Siu-Yung Jong¹, To Chan¹, Ming-Tak Hue² and Vincent W. L. Tam³

¹Department of Curriculum and Instruction & Centre for Learning Sciences and Technologies, The Chinese University of Hong Kong, Hong Kong // ²Department of Special Education and Counselling, The Education University of Hong Kong, Hong Kong // ³Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong // mjong@cuhk.edu.hk // chanto@cuhk.edu.hk // mthue@eduhk.hk // vtam@eee.hku.hk

ABSTRACT

There has been increasing discussion among educators and researchers about harnessing the idea of gamification to enhance the current learning and teaching practices in school education. Leveraging the context-aware mobile technology and student-centred learning theories, we have developed a mobile application, Gamified Authentic Mobile Enquiry in Society (GAMES), to support students in conducting authentic outdoor enquiry-based learning in social humanities education. This paper reports the quasi-experimental study in which we evaluated the learning effectiveness of GAMES in terms of supporting students’ knowledge construction, in comparison with the conventional outdoor enquiry-based learning approach. It involved a total of 559 Grade-10 students from top, middle, and bottom academic-banding schools. Results indicated that, compared to the conventional approach, GAMES had different degrees of positive effects on the high, moderate, and low academic-achieving participants. This study not only provides grounds for a wider adoption of GAMES in social humanities education, but also makes a contribution to empirical evidence in the field by designing, implementing and evaluating gamified mobile learning in authentic outdoor contexts.

Keywords

Social enquiry-based learning, Gamified learning, Authentic outdoor mobile learning, Context-aware technology

Introduction

There has been a persistent belief that learning in game-like contexts can promote students’ motivation and engagement (Games & Squire, 2011; Gee, 2013; O’Feil, Wainess & Baker, 2005; Papert, 1993; Piaget, 1970; Prensky, 2016). By taking advantage of this humans’ psychological predisposition, “gamification of learning” aims to integrate game mechanics and technologies into non-game learning environments to motivate and engage students in the pedagogic process (Burke, 2014; Kapp, 2012; Landers, 2014; Lee & Hammer, 2011). The recent K-12 edition of the New Media Consortium Horizon Report (Johnson, Adams Becker, Estrada & Freeman, 2016) forecasts that gamified learning will become a part of students’ lives in schools in the coming decade.

Rather than being passive knowledge recipients in traditional schooling, the twenty-first century education encourages students to play the active, learner-centric role to construct new knowledge on their own (Howland, Jonassen & Marra, 2012). Enquiry-based learning (EBL) has been one of the student-centred instructional practices advocated in today’s school education (Elder & Paul, 2009; Ho, 2012; Small, Arone, Stripling & Berger, 2012; Wallace & Husid, 2011). EBL requires students to construct knowledge self-directedly via recursive exploration and reflection in the learning process (Hwang, Chiu & Chen, 2015; Shih, Chuang & Hwang, 2010). While EBL can be applied in both science education and social humanities education, this paper focuses on the latter.

Traditional classrooms in schools are never a desirable venue for conducting EBL in social humanities education (Ip & Fok, 2010; Lim, 2004). To pursue meaningful explorative and reflective tasks in EBL, students are best to be situated in real-world, real-life contexts (Hill, 1994; Jansen, 2011; Small et al., 2012). Yet, the conventional type of outdoor fieldtrips is ineffective to engage and scaffold students in the course of EBL (Shih et al., 2010; Wake & Wasson, 2011; Zurita & Baloian, 2012; Jong, 2013, 2015b). While designing effective outside-the-classroom EBL activities remains a challenge in social humanities education (Johnson et al., 2016; Hwang et al., 2015), our work aims to address this pressing need.

Harnessing the context-aware mobile technology (in particular, the Global Positioning System [GPS]) with the basis on student-centred learning theories, we have developed a mobile application, Gamified Authentic Mobile Enquiry in Society (GAMES), to authentically support students in conducting outdoor EBL in social humanities education students in the pedagogic process (Burke, 2014; Kapp, 2012; Landers, 2014; Lee & Hammer, 2011). There has been increasing discussion among educators and researchers about harnessing the idea of gamification to enhance the current learning and teaching practices in school education. Leveraging the context-aware mobile technology and student-centred learning theories, we have developed a mobile application, Gamified Authentic Mobile Enquiry in Society (GAMES), to support students in conducting authentic outdoor enquiry-based learning in social humanities education. This paper reports the quasi-experimental study in which we evaluated the learning effectiveness of GAMES in terms of supporting students’ knowledge construction, in comparison with the conventional outdoor enquiry-based learning approach. It involved a total of 559 Grade-10 students from top, middle, and bottom academic-banding schools. Results indicated that, compared to the conventional approach, GAMES had different degrees of positive effects on the high, moderate, and low academic-achieving participants. This study not only provides grounds for a wider adoption of GAMES in social humanities education, but also makes a contribution to empirical evidence in the field by designing, implementing and evaluating gamified mobile learning in authentic outdoor contexts.

Keywords

Social enquiry-based learning, Gamified learning, Authentic outdoor mobile learning, Context-aware technology

Introduction

There has been a persistent belief that learning in game-like contexts can promote students’ motivation and engagement (Games & Squire, 2011; Gee, 2013; O’Feil, Wainess & Baker, 2005; Papert, 1993; Piaget, 1970; Prensky, 2016). By taking advantage of this humans’ psychological predisposition, “gamification of learning” aims to integrate game mechanics and technologies into non-game learning environments to motivate and engage students in the pedagogic process (Burke, 2014; Kapp, 2012; Landers, 2014; Lee & Hammer, 2011). The recent K-12 edition of the New Media Consortium Horizon Report (Johnson, Adams Becker, Estrada & Freeman, 2016) forecasts that gamified learning will become a part of students’ lives in schools in the coming decade.

Rather than being passive knowledge recipients in traditional schooling, the twenty-first century education encourages students to play the active, learner-centric role to construct new knowledge on their own (Howland, Jonassen & Marra, 2012). Enquiry-based learning (EBL) has been one of the student-centred instructional practices advocated in today’s school education (Elder & Paul, 2009; Ho, 2012; Small, Arone, Stripling & Berger, 2012; Wallace & Husid, 2011). EBL requires students to construct knowledge self-directedly via recursive exploration and reflection in the learning process (Hwang, Chiu & Chen, 2015; Shih, Chuang & Hwang, 2010). While EBL can be applied in both science education and social humanities education, this paper focuses on the latter.

Traditional classrooms in schools are never a desirable venue for conducting EBL in social humanities education (Ip & Fok, 2010; Lim, 2004). To pursue meaningful explorative and reflective tasks in EBL, students are best to be situated in real-world, real-life contexts (Hill, 1994; Jansen, 2011; Small et al., 2012). Yet, the conventional type of outdoor fieldtrips is ineffective to engage and scaffold students in the course of EBL (Shih et al., 2010; Wake & Wasson, 2011; Zurita & Baloian, 2012; Jong, 2013, 2015b). While designing effective outside-the-classroom EBL activities remains a challenge in social humanities education (Johnson et al., 2016; Hwang et al., 2015), our work aims to address this pressing need.

Harnessing the context-aware mobile technology (in particular, the Global Positioning System [GPS]) with the basis on student-centred learning theories, we have developed a mobile application, Gamified Authentic Mobile Enquiry in Society (GAMES), to authentically support students in conducting outdoor EBL in social humanities education students in the pedagogic process (Burke, 2014; Kapp, 2012; Landers, 2014; Lee & Hammer, 2011).
education. This paper reports the quasi-experimental study (with a total of 559 Grade-10 students) in which we evaluated the learning effectiveness of GAMES. Specifically, the study focused on probing into the effects of GAMES on supporting different academic-achieving students’ knowledge construction, in comparison with the conventional outdoor EBL approach.

We organise the rest of this paper as follows. The next section is a review of the related work. Then, the design and implementation of GAMES will be elaborated. After that, we will delineate the method, findings, implications, and limitations of the study. At the end of the paper, a conclusion of our observations in this study will be drawn.

Related work

Gamification and gamified learning

Video game-based learning or edutainment is about “the marriage of video games and education” (Prensky, 2016). Fun is always the best driving force for learning; “making learners feeling fun” by engaging them via gaming is a desirable motivating approach to education (Papert, 1993; Piaget, 1970). Video games (hereinafter referred to as games) are interactive activities with continuous challenges that engage players in an active learning process to master the rules and pursue the tasks therein (Koster, 2005). In fact, since the 1980s, there has been positive research evidence showing that harnessing games in the course of learning and teaching can effectively promote students’ motivation for eventually achieving the intended educative goals (e.g., Adam, 1998; Chee, 2016; Cordova & Lepper, 1996; Dede, 2011; Erhel & Jamet, 2013; Jong, 2015a; Lan, 2015; Malone, 1981; Shaffer, 2007). Instead of the direct adoption of games, an alternative way to leverage the idea of gaming in education is gamification. This term was selected as a runner-up for the word of the year by Oxford Dictionaries in 2011 (Burke, 2014).

The conception of “gamification” was first introduced in 2010 and initially adopted in the marketing field (Landers, 2014; Simões, Redondo & Vilas, 2013). After that, its application has been extended to other fields, including education. In Deterding, Dixon, Khaled and Nacke’s (2011) definition, “gamification is the use of game design elements and game mechanics in non-game contexts” (p. 9). In Kapp’s (2012) definition, “[g]amification is the use of game-play mechanics for non-game applications (also known as ‘funware’), particularly consumer-oriented web and mobile sites, in order to encourage people to adopt the application” (p. 10). In fact, so far there has been no consented definition of gamification; however, most of the definitions found in the literature do share similar features. Burke (2014) has given a concisely summarized elaboration on gamification — a strategy of using game mechanics and experience design to digitally engage people to achieve intended goals in non-game contexts. “Game mechanics” is about the core elements generically found in many games, e.g., points, badges, and leader boards. “Experience design” is about the experiential journey that engages players in a game, e.g., game play, play space, and story line. “Digitally” is about players’ interactions with computers, smartphones, tablets, or other digital devices. “Intended goals” are the pre-set goals that the organiser (e.g., a person, an organization, a company, etc.) wants players to reach at the end of a gamified activity.

A common and key challenge faced by schools for years has been still around students’ motivation in learning (Lee & Hammer, 2011). The initiative of gamification in education attempts to use game-like environments and game-play experiences to engage students in educational activities. Strategies of educational gamification are generally termed gamified learning (Kapp, 2012; Landers, 2014; Simões et al., 2013). Again, the field has yet to come up with a single, consented definition for gamified learning. In this paper, on top of Burke’s (2014) elaboration on gamification, we adopt Domínguez’s et al. (2014) definition of gamified learning — an approach to integrating the idea of gamification into the learning process to support students in attaining intended educative goals engagingly.

To gamify the course of learning, it is not just a matter of infusing how many game mechanics into a learning environment (Landers, 2014). The focus should be placed on the meaningful integration of the mechanics into experience design for sustaining students’ continuous participation in the environment (Domínguez et al., 2013; Kapp, 2012). Lee and Hammer (2011) have proposed a design framework for developing gamified learning environments. This framework, which guides our present work, contains three vital components:

- The cognitive component. A gamified learning environment should contain a system of rules articulated to a series of learning tasks that lead students to master the rules. Each task (and the corresponding goal) can be further divided into a number of sub-tasks (and the corresponding sub-goals) to lower students’ cognitive
load (Sweller, Ayres & Kalyuga, 2011) in the learning process. In addition, to better support students in attaining all of the goals, it is desirable to offer them some non-linear flexibility to complete the tasks in accordance with their own ability and personal preference, without forcing them to follow a specific sequence.

- The emotional component. A gamified learning environment should be able to provide students with feelings of “success” and “failure” in the learning process. It is expected that students will have positive emotions when they have successfully accomplished a learning task. To stimulate their positive emotions, the reward system (underpinning to the environment) should offer them some sorts of immediate awards, such as points, badges, levelling-up the progress bar, unlocking the next task, etc. On the contrary, students will have some degree of anxiety when they have failed to accomplish a learning task. However, to avoid turning their anxiety into frustration, it is desirable to keep the stakes low and the penalty mild, so that they are still willing to actively participate in the learning environment.

- The social component. A gamified learning environment should possess learning tasks that facilitate social interactions among students. The interactions can be in the form of collaboration (e.g., working towards a common goal) and/ or competition (e.g., contending to perform better than others). Both collaborative and competitive interactions will provoke students’ new “public” identities in the environment. These identities will bring them the new social credibility and recognition, which will have a beneficial impact on learning, such as enhancing student engagement in pursuing the tasks and initiating new approaches to their learning participation.

There has been research evidence showing that gamified learning can effectively strengthen students’ learning motivation (similar to the findings in game-based learning research in general), but may not consequentially promote their learning performance (Attali & Arieli-Attali, 2015; Dominguez et al., 2014; Muntean, 2011; Simões et al., 2013). In fact, gamified learning should not solely be aimed at increasing students’ willingness to participate in learning, but also providing them with a more supportive environment that favours a higher level of knowledge construction (Burke, 2014; Howland et al., 2012; Kapp, 2012). Pedagogic integration of learner-centric, constructivist learning theories into the design of gamified learning for promoting deep learning is an important area that needs more research effort (Landers, 2014; Lee & Hammer, 2011).

Enquiry-based learning (EBL)

Knowledge in all disciplines is continuously expanding (Chee, 2016; Gee, 2013; Howland et al., 2012). Therefore, equipping today’s youngsters with the ability to become “knowledge builders” and “life-long learners” is far more vital than to master the contents in textbooks. EBL (also named “inquiry-based learning” in the literature) emphasises the development of students’ skills of and dispositions to knowledge construction and life-long learning (Wallace & Husid, 2011). Both the United Nations Educational, Scientific and Cultural Organisation (UNESCO) and the Organisation for Economic Co-operation and Development (OECD) have regarded EBL as one of the important instructional approaches to facilitating meaningful student-centred learning in school education (Ho, 2012).

In Wallace and Husid’s (2011) definition, EBL is the process of “seeking knowledge, raising questions, searching for answers, evaluating information, and asking new questions based on new understandings” (p. 21). In Small’s et al. (2012) definition, EBL is a process that “involves connecting to personal interests and a desire to know, gaining background knowledge, asking questions that probe beyond simple fact gathering, investigating answers to gather evidence from multiple perspectives and sources, constructing new understandings and drawing conclusions with support from evidence” (p. 3). In fact, there had been no standard definition of EBL among educators and researchers. Nevertheless, most of EBL strategies in the literature (e.g., Chadwick, 2008; Eisenberg, Berkowitz & Johnson, 2010; de Jong, 2016; Jansen, 2011; Stripling, 2003; Stripling, 2008) do share common pedagogic aims and involve similar enquiry actions, such as raising questions, searching for knowledge and information to answer the questions, and asking new questions based on continuous reflection (Elder & Paul, 2009).

The application of EBL can be found in both science education (e.g., de Jong, 2006; Ucar & Trundle, 2011) and social humanities education (e.g., Hwang et al., 2015; Lim, 2004; Shih et al., 2010). Different from scientific EBL that focuses on probing into the physical truths on the “natural” earth, social EBL emphasizes looking into humans and their relationships with the “societal” world from multiple perspectives, values and interests (Hill, 1994; Jansen, 2011; Small et al., 2012). In social humanities education, EBL usually pivots on a societal issue which is of real-life, open-ended, complex, changing, and/ or controversial nature (Chadwick, 2008; Small et al., 2012), e.g., “to what extent are traditional customs compatible with modern society.” Inspired by Dewey’s
...theoretical framework, through the EduVenture® Composer (a cloud-based authoring system, see http://ev-cuhk.net) (Jong & Tsai, 2016), we have developed GAMES—a tablet-based GPS-supported mobile application (viz., an App) to gamify and mobilise social EBL in authentic outdoor environments. With GAMES, during the course of enquiry, students can plan their own route and control the time to be spent at each exploratory spot according to their own ability and interest, without largely relying on their teachers.

GAMES uses a two-tier interface approach to presenting the digital material that supports students in probing into a specific societal issue in an outdoor environment (Jong & Tsai, 2016). The first-tier interface is an...
an electronic map of the fieldtrip site (see Figure 1). Students can select either the graphic view (Figure 1a) or the satellite view (Figure 1b) based on their own preference. An avatar on the map is used to indicate the authentic geographical position of each student at the site (see the avatar of a student, John, in Figure 1). Based on the GPS information continuously received by the tablet, John’s avatar on the screen will move in accordance with his physical movement in the fieldtrip. He can also see the movements of other avatars that denote his classmates in a real-time manner. A number of exploratory spots (with each embedded with a specific enquiry task) are pre-marked on the map (see the “stop signs” in Figure 1). The task at each spot has yet to be accessible until John physically arrives at the corresponding geo-location at the site.

![Figure 1. First-tier interface of GAMES](image)

![Figure 2. Checking-in an exploratory spot and unlocking the task therein](image)
A location which is selected as an exploratory spot in GAMES should possess (i) high contextual relatedness to the society issue, and (ii) high affordance for embedding the four enquiry sub-tasks (i.e., Connection, Investigation, Construction and Reflection) framed by Stripling's (2003, 2008) EBL model. Let us continue to use John as an example to illustrate how GAMES supports a student in outdoor EBL. After John has checked in the spot, the task will be unlocked and the “stop sign” on the screen will become a “blue pin” (see Figure 2). Then, by tapping on the pin, he can enter the second-tier interface (see Figure 3) to access the task which is composed of four sub-tasks (viz. digital scaffolds):

- The Connection sub-task/scaffold, which is in the form of voice navigation, equips John with the background information about the issue and relate his prior knowledge to the information. Figure 3a shows a voice-navigation Connection scaffold.

- The Investigation sub-task/scaffold, which is in the form of data-collection exercise (e.g., context capturing via photo-taking and video-recording gadgets, environmental surveys, as well as audio-recording interviews with the locals), guides John to gather new information to answer the questions related to the issue. Figure 3b shows a context-capturing Investigation scaffold.

- The Construction sub-task/scaffold requires John to work with two to three neighbourhood classmates currently at the spot to discuss and generalise the information that they have individually gathered in the Investigation sub-task via mind mapping, as well as proposing an interim conclusion about the issue with the evidence on hand via audio recording. Figure 3c shows a mind-mapping Construction scaffold.
The Reflection sub-task/scaffold assists John, via video blogging, in reflecting on his weaknesses when exploring this spot (e.g., too shy to interview the locals) and the limitations of the interim conclusion, as well as setting new goals for exploring the next spot. Figure 3d shows a video-blogging Reflection scaffold.

After John has completed all the four sub-tasks at an exploratory spot, he will be immediately awarded a “star” on his progress bar (see Figure 4). On the other hand, if he leaves a spot without completing all the sub-tasks therein, the App will vibrate to alert and request him to revisit the spot for finishing the missed sub-task(s). A leader board in the App will dynamically indicate how many stars each classmate has obtained and how much time he/she has spent on obtaining the stars in a real-time manner. Figure 5 captures the upper part of the leaderboard showing that John is currently on the top because so far he has got the largest number of stars with the shortest time. Table 1 summarizes the game mechanics and experience design that are integrated into GAMES to support the course of outdoor EBL in accordance with Lee and Hammer’s (2011) gamified learning framework in the cognitive, emotional and social aspects.

<table>
<thead>
<tr>
<th>Aspect of gamified learning (Lee &amp; Hammer, 2011)</th>
<th>Game mechanics and experience design integrated into GAMES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive</td>
<td>• According to his own preference, John can flexibly plan his enquiry route and control his time to be spent at each exploratory spot according to his ability and interest, without being “forced” to follow a specific sequence.</td>
</tr>
<tr>
<td></td>
<td>• The exploratory task (goal) at each spot is divided into sub-tasks (sub-goals) to lower John’s cognitive load in the learning process.</td>
</tr>
<tr>
<td></td>
<td>• The design of the sub-tasks is ruled by Stripling’s (2003, 2008) EBL model; in other words, the rules of Connection, Investigation, Construction, and Reflection guide John to probe into every spot to attain the corresponding goal.</td>
</tr>
<tr>
<td>Emotional</td>
<td>• When John successfully locates an exploratory spot (via his physical movement at the fieldtrip site), he can unlock the exploratory task therein.</td>
</tr>
<tr>
<td></td>
<td>• A “star” (visible achievement) will be immediately awarded to John and shown on his progress bar once he has successfully completed all the sub-tasks at a spot.</td>
</tr>
<tr>
<td></td>
<td>• If John has failed to complete all the sub-tasks before leaving a spot, apart from not being awarded a star, he will be alerted and requested to go back to</td>
</tr>
</tbody>
</table>
the spot to finish the uncompleted sub-task(s). This low-stakes penalty aims to maintain John’s willingness to go on participating in the fieldtrip.

Social
- To eventually get a star after visiting an exploratory spot, when conducting the Construction sub-task, John has to work collaboratively with his classmates to discuss and generalise the collected information to give an interim conclusion about the issue with evidence.
- During the fieldtrip, John is competing for an upper position on the leaderboard against other classmates, based on how many stars that he has obtained and how much time he has spent so far to get the stars.
- Both collaborative and competitive interactions provoked in the learning process stimulate John to develop a new “public” identity in the learning context and make him more engaged in the fieldtrip.

Research design

We employed a quantitative approach to answering the research question of this study: “Can the adoption of GAMES promote different academic-achieving students’ knowledge construction in social humanities education, in comparison with the conventional outdoor EBL approach?” The collection of qualitative data was to serve the purpose of supplementing the quantitative findings.

Liberal Studies (LS)

Liberal Studies (LS) is a “young” core subject of social humanities education in the new senior secondary education system (i.e., Grade-10 to Grade-12) under the recent education reform in Hong Kong (Education Bureau, 2014). The Hong Kong Diploma of Secondary Education Examination (HKDSE), which is a high-stakes public examination, has started to include LS as an examination subject since 2012 (Curriculum Development Council, 2014). The core curricular aim of LS is to develop students’ knowledge and multiple perspectives on various societal issues in cultural, social, economic, political and technological contexts.

The curriculum of LS is composed of a number of thematic areas. Each area consists of different enquiry modules, and every module consists of a number of societal issues. For example, the issue involved in this study, “to what extent are traditional customs compatible with modern society?” is under the “Culture and Modern Life” module which belongs to the “Society and the Environment” area. Stripling’s (2003, 2008) EBL model is one of the popular instructional practices adopted by the LS teachers in Hong Kong.

Participating schools, teachers and students

Secondary schools in Hong Kong are categorised into three academic bands based on the overall academic performance of their students in attaining the learning objectives of the formal school curriculum; Band A, Band B, and Band C are respectively the top, middle, and bottom bands. We selected three schools at each academic banding among our research partners to take part in this study (i.e., 9 schools in total; 3 Band-A schools, 3 Band-B schools, and 3 Band-C schools). The following were the selection criteria:
- Each school should provide two Grade-10 classes to participate in the study.
- There is no significant difference ($p > .05$) between the LS examination mean scores of these two classes in the previous semester.
- The classes have yet to learn the “Culture and Modern Life” module.
- The classes have experience in participating in conventional outdoor fieldtrips (with paper-based worksheets) framed by Stripling’s (2003; 2008) EBL model in the previous semester.
- The LS teachers among the selected schools have comparable academic background and years of experience in facilitating conventional outdoor fieldtrips (with paper-based worksheets) framed by Stripling’s (2003, 2008) EBL model.

The nine teachers respectively from the nine selected schools each possessed a master degree and around five years of the related outdoor EBL facilitation experience. Among the nine schools, a total of 559 students participated in the study; 192 were high-achieving students (from the Band-A schools), 190 were moderate-achieving students (from the Band-B schools), and 177 were low-achieving students (from the Band-C schools).
Their average age was 16.35. In each school, one class was assigned to the experimental group, and another class was assigned to the control group. Table 2 shows the distribution of different academic-achieving students in the experimental and control groups.

<table>
<thead>
<tr>
<th></th>
<th>High-achieving students</th>
<th>Moderate-achieving students</th>
<th>Low-achieving students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control groups</td>
<td>95</td>
<td>95</td>
<td>88</td>
</tr>
<tr>
<td>Experimental groups</td>
<td>97</td>
<td>95</td>
<td>89</td>
</tr>
</tbody>
</table>

**Experimental procedure and data collection**

The quasi-experiment of this study involved enquiring into the societal issue, “to what extent are traditional customs compatible with modern society.” The experimental manipulation was outdoor EBL with GAMES, while the control manipulation was outdoor EBL with paper-based worksheets (i.e., the conventional approach). For writing convenience, we use G-EBL and P-EBL to denote the experimental manipulation and the control manipulation respectively.

Six months before the experiment, we started designing for the fieldtrip and the supporting materials for both G-EBL and P-EBL. The preparation included determining the fieldtrip site, exploratory spots and duration, as well as developing the contents to be used in GAMES for G-EBL and the worksheets for P-EBL. The design of the contents for both manipulations was based on Stripling’s (2003; 2008) EBL model. Each contained the same number of exploratory spots, and the task at each spot was composed of the same sub-tasks of Connection, Investigation, Construction, and Reflection (learning scaffolds). The main difference was that the content of GAMES was in a gamified manner and G-EBL students responded to the learning scaffolds digitally (as described previously in the “GAMES” section), while the content of the worksheets was in a conventional text-based manner and P-EBL students responded to the learning scaffolds with written text.

Four months before the experiment, we set up a review committee to evaluate the materials developed for G-EBL and P-EBL to ensure that they aligned with the curricular aim of the “Culture and Modern Life” module and were of comparable quality. The committee was composed of two education professors respectively from two other institutions, six LS teachers respectively from six non-participating schools (2 at Band A, 2 at Band B, and at 2 at Band C), and one government curriculum officer. Two months before the experiment, we explained to each of the nine teachers the research details, as well as familiarising him/her with the pedagogic idea of G-EBL and the technical operation of GAMES.

The same experiment was separately carried out in each participating school. The size of each group (experimental/control) ranged from 29 to 33. The same teacher in each school facilitated both the experimental manipulation (G-EBL) and control manipulation (P-EBL). The manipulation and the corresponding data collection work conducted in each group were completed within three consecutive days, involving the following procedures:

- **Day 1: Briefing.** One day before the fieldtrip, the teacher conducted a 35-minute briefing session to tell each group about the arrangement of the fieldtrip, such as the location, transportation, duration, assembling/dismissing points, safety issues, and how to use the fieldtrip material (GAMES/ worksheets).
- **Day 2: Fieldtrip.** The fieldtrip involved five hours in total. Without the teacher’s intervention, the experimental group conducted G-EBL with the digital supporting material (with GAMES). Led by the teacher in a conventional manner (as aforementioned in the “Related work” section), the control group conducted P-EBL with the paper-based supporting material (with the worksheets).
- **Day 3 morning: Knowledge test.** One day after the fieldtrip, we administered an unseen knowledge test (one hour) to each group. The test was in Chinese and in the typical format of the LS public examination in Hong Kong. It consisted of three open-ended questions which were customised from the recent five-year public examination questions related to the societal issue used in this study. The perfect score of the test was 45 (15 marks for each question). The validity of the test and marking scheme were scrutinised by the review committee (as aforementioned in the previous sub-section). All completed test papers were individually marked by three trained markers who were postgraduate education students majoring in LS in the first author’s university. The marking was done anonymously, i.e., the school and group information on the test papers were removed before they were passed to the markers. The first author was responsible for reconciling and discerning the discrepancies among the markers’ work.

<table>
<thead>
<tr>
<th>Table 2. Distribution of different academic-achieving students in the experimental and control groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control groups</td>
</tr>
<tr>
<td>High-achieving students (n = 192)</td>
</tr>
<tr>
<td>Control groups</td>
</tr>
<tr>
<td>Experimental groups</td>
</tr>
</tbody>
</table>

285
• Day 3 afternoon: Interview. Right after the test, we randomly selected two students (one male and one female) from the experimental group for a group interview (45 minutes). We aimed to gain more understanding about how they perceived G-EBL. The interview was audio-recorded, and conducted in Chinese (the mother tongue of the students) and in semi-structured format. The guiding question was “How is your learning experience with GAMES, in comparison with your past fieldtrip learning experience?.” We transcribed the interview clips and adopted Maxwell’s (2013) four-phase qualitative analysis approach (coding, categorizing, memoing, and contextualizing) to analyzing the interview transcript and complement the knowledge test results.

Results

We received a total of 542 completed knowledge test papers from the nine schools (return rate = 96.96); 185 were completed by the high-achieving students, 182 by the moderate-achieving students, and 175 by the low-achieving students. The findings regarding different kinds of academic-achieving students are presented in the following sub-sections.

High-achieving students

Table 3 shows the descriptive statistics of the test scores obtained respectively from the experimental and control groups in the Band-A schools. An independent samples t-test on the scores indicated that the experimental-group students’ average score (35.55) was significantly different from the control-group students’ (32.81), t(183) = 1.88 p < .05). The Cohen’s d was 0.28. In other words, the learning effect of G-EBL on supporting the high-achieving students’ knowledge construction was stronger than P-EBL’s, but the effect size was small (Cohen, 1998). We further conducted an independent samples t-test to assess if the gender factor (male students vs female students) had an influence on the test results obtained from the experimental group. The analyses indicated that there was no significant difference (p > .05).

<table>
<thead>
<tr>
<th></th>
<th>Experimental groups (n = 92)</th>
<th>Control groups (n = 93)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>35.55</td>
<td>32.81</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>10.03</td>
<td>9.86</td>
</tr>
</tbody>
</table>

When we interviewed the experimental-group students, they shared their desirable learning experience with GAMES, in comparison with the conventional outdoor EBL approach that they had experienced before. The following are some translated interview excerpts:

• Aaron (pseudonym): In the conventional fieldtrip, we used to write a lot on the worksheets …… without a proper table, I cannot write properly … it is also time-consuming … but yesterday, with the App, I could use photos, audios and videos to document most of my observations during the exploratory process.

• Amy (pseudonym): The approaches to checking-in the exploratory spots and unlocking the enquiry tasks were quite novel to me. It was like playing a treasure hunt game … it is much funnier than what we used to do in the conventional fieldtrip … I felt happy whenever I successfully located a spot at the fieldtrip site.

• Andrew (pseudonym): In the conventional fieldtrip, the guidance on the worksheets is largely text-based … I found the voice navigation function of the App very useful … I could listen to the guidance while making the observation accordingly.

• Andy (pseudonym): Undoubtedly, I did not want to be at the bottom position (on the leader board), haha … otherwise, I would be so ashamed … Yes, it was more interesting than the conventional fieldtrip … and the competitive atmosphere did make me more engaged in the learning process.

• Angel (pseudonym): Unlike the conventional fieldtrip, you will never miss any sub-tasks when using this intelligent App. It will alert and request you to finish the sub-tasks even though you have left the spot … in fact, this function did help me twice during the fieldtrip yesterday … thanks for the App.

• Ann (pseudonym): Yesterday, I could decide my own exploratory route during the fieldtrip. It made the whole learning process more self-directed … Also, the collaboration in the Construction sub-task at each spot provided me with more insights into the societal issue from multiple angles. This never happens in the conventional fieldtrip. I hope my teacher can give us more chances to use this App to conduct outdoor learning in LS in the future.
Moderate-achieving students

Table 4 shows the descriptive statistics of the test scores obtained respectively from the experimental and control groups in the Band-B schools. An independent samples t-test on the scores indicated that the experimental-group students’ average score (29.86) was significantly different from the control-group students’ (22.65), t(180) = 4.61, p < .001. The Cohen’s d was 0.69. In other words, the learning effect of G-EBL on supporting the moderate-achieving students’ knowledge construction was stronger than P-EBL’s. The effect size was medium to large (Cohen, 1998). We further conducted an independent samples t-test to assess if the gender factor (male students vs female students) had an influence on the test results obtained from the experimental group. The analyses indicated that there was no significant difference (p > .05).

<table>
<thead>
<tr>
<th></th>
<th>Experimental groups (n = 92)</th>
<th>Control groups (n = 90)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>29.84</td>
<td>22.65</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>10.12</td>
<td>10.64</td>
</tr>
</tbody>
</table>

When we interviewed the experimental-group students, they elaborated that they were more engaged in the course of learning with GAMES, in comparison with the conventional outdoor EBL approach that they had experienced before. The following are some translated interview excerpts:

- **Ben (pseudonym):** It was so convenient to learn with this App … I could quickly gather the information that I observed during the fieldtrip through photo taking and audio recording. In the conventional fieldtrip, I can only use text to document my observations … it is not an effective approach as sometimes it is very hard to record all details in words ….. fortunately, the App alerted me to revisit the spot … I had overlooked a sub-task at that spot … I got back the star at that spot finally, haha.

- **Bethany (pseudonym):** Yesterday, it was really like playing a game … in the activity, I wanted to get all stars and obtain the No. 1 position on the leader board ….. it was not just about competition, but also collaboration. I did work together with the classmates when doing the Construction sub-task at each spot. The competitive and collaborative atmosphere made me very engaged in the whole learning process.

- **Betty (pseudonym):** The process of finding out the exploratory spots in the activity was so interesting, like Facebook™ check-in ….. I never have this sense of success when participating in the conventional fieldtrip. It was also funny to see other classmates (the avatars) who took different exploratory routes moving on the App.

- **Billy (pseudonym):** I deem that doing reflection via video blogging is a very efficient way, especially when we are learning outside the classroom. In fact, using text to document the reflection in the conventional fieldtrip is very time consuming and inconvenient ….. I prefer doing more explorative work rather than writing during a fieldtrip.

- **Bobby (pseudonym):** The App provided us with both collaborative and competitive tasks. Yesterday, on the one hand, we had to work together at each spot to co-construct the mind map. On the other hand, we were contending for the upper positions on the leader board. Both collaborative and competitive atmospheres made the whole learning process very engaging … especially the competition element which will never be found in the conventional fieldtrip.

- **Brenda (pseudonym):** The App allowed me to decide my exploratory route and time spent at each spot, without relying much on the teacher ….. Different from the conventional fieldtrip, yesterday the whole class did not crowd into a spot simultaneously. Therefore, we had more room to enquire about the context of every spot and interact with the local people around.

Low-achieving students

Table 5 shows the descriptive statistics of the test scores obtained respectively from the experimental and control groups in the Band-C schools. An independent samples t-test on the scores indicated that the experimental-group students’ average score (20.07) was significantly different from the control-group students’ (12.95), t(173) = 6.33, p < .001. The Cohen’s d was 0.81. In other words, the learning effect of G-EBL on supporting the low-achieving students’ knowledge construction was stronger than P-EBL’s. The effect size was large (Cohen, 1998). We further conducted an independent samples t-test to assess if the gender factor (male students vs female students) had an influence on the test results obtained from the experimental group. The analyses indicated that there was no significant difference (p > .05).
When we interviewed the experimental-group students, they expressed that GAMES made the whole learning process more supportive and exciting, in comparison with the conventional outdoor EBL approach that they had experienced before. The following are some translated interview excerpts:

- **Calvin (pseudonym):** Unlike the conventional fieldtrip, this time I could audio-record the contents of the interviews with the locals. It helped me a lot because I am an absent-minded person. I will easily forget what interviewees said. Yesterday I could audio-record the content first and re-play the sound bites again when needed, such as when conducting the Construction sub-task and Reflection sub-task at each spot.

- **Chloe (pseudonym):** I like gaming … The yesterday activity was very much like a game. Everyone was like a player … I was so happy that I could decide my own exploratory route … I had never been so engaged in participating in a learning activity … but yesterday I completed all the learning tasks. After finishing all the sub-tasks at a spot, I was looking forward to accessing the next one …… Since I did not know what I would do at the next spot, it made the whole learning process very exciting.

- **Christine (pseudonym):** I hate to participate in the conventional fieldtrip because the tasks are usually very boring. However, yesterday I was quite engaged …… I didn’t miss any tasks as the App would always alert me if I missed anything …… I felt energetic when I was awarded a star after completing the task at a spot.

- **Colin (pseudonym):** I hate writing, so I love this App so much …… I can speak a lot in front of a camera. However, if you ask me to express all the things with text, I don’t think I am able to do it. So, I don’t like the conventional fieldtrip. This App is an excellent tool for students, like me, who are not good at writing, haha.

- **Connie (pseudonym):** The real-time updating feature of the leader board made the learning process very exciting. Although I was not very enthused about winning the No. 1 position, I did not want to be positioned at the bottom, haha …… I also made use of the information on the board to evaluate my on-going progress by comparing with other classmates’.

- **Conrad (pseudonym):** I am a super gamer, and I always win in video gaming, haha. That’s why I was so into this activity … I wanted to get the No. 1 position on the leader board. See, yesterday I did it, haha … I collaborated with my classmates to conduct the Construction sub-task at each spot seriously. Without using this App, I think I wouldn’t have this enthusiasm.

### Discussion

In this study, the learning scaffolds in the experimental and control manipulations were both framed by Stripling’s (2003; 2008) EBL model. The main difference was that the learning process of the experimental manipulation (G-EBL) was further modelled by Lee and Hammer’s (2011) gamified learning framework. The experimental-group students interacted with the learning scaffolds digitally with GAMES (c.f., the paper-based worksheets used in P-EBL). According to the knowledge test results, in comparison with the conventional approach, the G-EBL had desirable learning effects on the high-, moderate- and low-achieving participants, respectively with small, medium-to-large and large effect sizes. In the interviews with the experimental-group participants, no matter which academic banding they were from, they regarded that GAMES provided them with more engaging EBL experience, in comparison with the conventional fieldtrips that they had experienced before. The findings can be further elaborated from the cognitive, emotional and social perspectives in accordance with Lee and Hammer’s framework.

From the cognitive perspective, without being forced to follow a specific sequence of visiting the exploratory spots at the fieldtrip site, the students in G-EBL could set up their own self-planned enquiry journey according to their own ability and interest (e.g., see the interview excerpts of Ann, Brenda and Chloe). It mitigated the problems of “teacher-oriented learning” and “low learning motivation” in P-EBL (Jong, 2013; Jong, 2015b). Moreover, like players mastering “game rules” in gaming, the students in G-EBL were engaged in pursuing the exploratory tasks which were ruled by Stripling’s (2003; 2008) EBL model. The task at each spot presented in GAMES was further divided into the sub-tasks of Connection, Investigation, Construction, and Reflection. The digital gadgets (e.g., voice navigation, context capturing, video blogging) in GAMES also assisted them in conducting the sub-tasks more efficiently and effectively (e.g., see the interview excerpts of Aaron, Andrew, Ben, Billy, Calvin and Colin). All these “germane supports” (Sweller et al., 2011) reduced the students’ cognitive load in the knowledge construction process, tackling the problem of “non-engaging learning scaffolds” in P-EBL (Jong, 2013; Jong, 2015b).

<table>
<thead>
<tr>
<th></th>
<th>Experimental groups (n = 89)</th>
<th>Control groups (n = 86)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average</strong></td>
<td>20.07</td>
<td>12.95</td>
</tr>
<tr>
<td><strong>Standard deviation</strong></td>
<td>9.11</td>
<td>8.52</td>
</tr>
</tbody>
</table>

Table 5. Low-achieving students’ knowledge test results
From the emotional perspective, the students in G-EBL perceived both positive and negative emotions similar to what players experience in gaming. They had a feeling of success when they located an exploratory spot with GAMES at the fieldtrip site and unlocked the exploratory task therein, as well as getting a star immediately after completing all the sub-tasks at the spot (e.g., see the interview excerpts of Amy, Bethany, Betty, Chloe and Christine). On the other hand, they had a feeling of failure when GAMES alerted and requested them to revisit the spot to complete the unfinished sub-task(s). Although failure will usually induce frustration that obstructs learners’ continuous participation (Koster, 2005), the low-stakes penalty (revisiting alerts and requests) used in GAMES was able to keep the failed students in the “flow” state (Csikszentmihalyi, 2008) to go on their learning with enthusiasm (e.g., see the interview excerpts of Angel, Ben and Christine). As evidenced, GAMES provided the students with a more engaging knowledge construction environment, alleviating the problems of “low learning motivation” and “non-engaging learning scaffolds” in P-EBL (Jong, 2013; Jong, 2015b).

From the social perspective, the students in G-EBL were highly motivated to interact with one another both collaboratively and competitively. To obtain a star at each exploratory spot, they had to work toward the common goal of accomplishing the Construction sub-task (e.g., see the interview excerpts of Ann, Bethany, Bobby and Conrad). On the other hand, the students were also enthused to contend for the upper positions on the leader board in GAMES (e.g., see the interview excerpts of Andy, Bethany, Bobby, Connie and Conrad). It moderated the problem of “lack of peer interactions” in P-EBL (Jong. 2013; Jong, 2015b). Like gaming, both the collaborative and competitive interactions stimulated the students to build up new “social identities” (Lave & Wenger, 1991) during the fieldtrip. The credibility and recognition brought by the identities increased their willingness to devote their effort to construct new knowledge in the learning process (Scardamalia & Bereiter, 2006).

A possible limitation of this study might be the Hawthorne effect (McBride, 2015) on the experimental-group participants. Can the positive learning effects of G-EBL on these participants last? There is a need for further studying the substantiality/ change of the effects when they learn with GAMES again to enquire into different societal issues in other curricular modules. In addition, regarding the small effect size obtained in the top academic-achieving schools (though the experimental groups significantly outperformed the control groups), it might be owing to the ceiling effect (Cramer & Howitt, 2004). High-achieving learners generally have stronger cognitive ability and do well under conventional pedagogic practices in schools (e.g., P-EBL) (Biggs & Moore, 1993). Nevertheless, at this stage, we will not simply draw a quick conclusion that G-EBL is relatively less effective for academically-top learners. Instead, supported by the positive feedback gathered from the high-achieving students in the present research, we will further investigate whether GAMES can better empower academically-top learners when they explore more complex societal issues identified in the meta-studies on the LS curriculum (Fung, Tang & Chan, 2011; Ip & Fok, 2010), i.e., to moderate the ceiling effect by adjusting the cognitive load (Sweller et al., 2011) in the enquiry process.

Conclusion

In this paper, we have presented our educational gamification initiative that aims to gamify and mobilise an instructional practice, outdoor EBL, in the context of social humanities education. We have also reported and discussed the quasi-experimental study for evaluating the learning effectiveness of the initiative in secondary schools. Prior research has shown that gamified learning can generally promote students’ learning motivation, but not consequentially enhance students’ learning performance. However, this study offers new evidence for showing that well-designed pedagogic gamification can not only promote learners’ motivation but also significantly enhance their knowledge construction performance. It provides the field with grounds for further investigating a wider adoption of gamified learning in school education, making a contribution to designing, implementing and evaluating gamified mobile learning in authentic outdoor environments. Moreover, as aforesaid, LS is a “young” core social humanities subject in Hong Kong. Many teachers are still exploring the effective approaches to learning and teaching of this subject. The statutory curriculum document emphasises the importance of giving students the authentic outdoor learning opportunities to interact with the society. Our work offers local teachers, teacher educators, and education policymakers a practical reference for supporting outside-the-classroom learning and teaching activities in LS.

Acknowledgements

The work described in this paper was substantially supported by a grant from the Research Grants Council of Hong Kong Special Administration Region, China (Project No.: 14616815). We very much thank the
participating schools, teachers and students. In addition, we very much appreciate the scholarly and professional advice provided by the members of the ad hoc review committee formed for this study.

References


