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 The articles should be original, unpublished, and not in consideration for publication elsewhere at the time of submission to Educational Technology & Society and three months thereafter.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

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

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

Technical Manager
Panagiotis Zervas, University of Piraeus & Centre for Research and Technology Hellas, Greece

Executive Peer-Reviewers
http://www.ifets.info/

Publisher
International Forum of Educational Technology & Society

Editorial Office
c/o Dr Nian-Shing Chen, Department of Information Management, National Sun Yat-sen University, 70, Lien-Hai Rd, Kaohsiung, 80424, Taiwan.
Supporting Organizations

Center for Cognition and Gesture-based Computing, National Sun Yat-sen University, Taiwan
Centre for Research and Technology Hellas, Greece

Subscription Prices and Ordering Information

For subscription information, please contact the editors at ets-editors@ifets.info.

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@ifets.info.

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:
http://www.ifets.info/guide.php

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.

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.

The submissions should be uploaded at http://www.ifets.info/ets_journal/upload.php. In case of difficulties, please contact ets-editors@ifets.info (Subject: Submission for Educational Technology & Society journal).
# Table of contents

## Full Length Articles

<table>
<thead>
<tr>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Effects of Captioning Videos on Academic Achievement and Motivation: Reconsideration of Redundancy Principle in Instructional Videos</td>
<td>1–10</td>
</tr>
<tr>
<td>Mazaffer Ozdemir, Serkan Izmirli and Ozden Sahin-Izmirli</td>
<td></td>
</tr>
<tr>
<td>Computer Card Games in Computer Science Education: A 10-Year Review</td>
<td>11–21</td>
</tr>
<tr>
<td>Maria Kordaki and Anthi Gousiou</td>
<td></td>
</tr>
<tr>
<td>ePortfolio-Based Learning Environments: Recommendations for Effective Scaffolding of Reflective Thinking in Higher Education</td>
<td>22–33</td>
</tr>
<tr>
<td>Pauline Roberts, Dorit Maar and Jan Herrington</td>
<td></td>
</tr>
<tr>
<td>An Analysis of Density and Degree-Centrality According to the Social Networking Structure Formed in an Online Learning Environment</td>
<td>34–46</td>
</tr>
<tr>
<td>Esin Ergin and Yasemin Koçak Usluel</td>
<td></td>
</tr>
<tr>
<td>Reading Subtitles and Taking Enotes While Learning Scientific Materials in a Multimedia Environment: Cognitive Load Perspectives on EFL Students</td>
<td>47–58</td>
</tr>
<tr>
<td>John J. H. Lin, Yuan-Husan Lee, Dai-Yi Wang and Sunny S. J. Lin</td>
<td></td>
</tr>
<tr>
<td>Comparing the Effectiveness of Self-Learning Java Workshops with Traditional Classrooms</td>
<td>59–74</td>
</tr>
<tr>
<td>Eranki L. N. Kiran and Kannan M. Moudgalya</td>
<td></td>
</tr>
<tr>
<td>Effects of WOE Presentation Types Used in Pre-training on the Cognitive Load and Comprehension of Content in Animation-Based Learning Environments</td>
<td>75–86</td>
</tr>
<tr>
<td>Jaewon Jung, Dongyuk Kim and Chungsoo Na</td>
<td></td>
</tr>
<tr>
<td>Which Teaching Strategy is Better for Enhancing Anti-Phishing Learning Motivation and Achievement?</td>
<td>87–99</td>
</tr>
<tr>
<td>The Concept Maps on Tablet PCs or Worksheets?</td>
<td></td>
</tr>
<tr>
<td>Jerry Chih-Yuan Sun and Kuan-Hsien Lee</td>
<td></td>
</tr>
<tr>
<td>Effects of A Peer Assessment System based on a Grid-based Knowledge Classification Approach on Computer Skills Training</td>
<td>100–111</td>
</tr>
<tr>
<td>Ting-Chia Hsu</td>
<td></td>
</tr>
<tr>
<td>Design and Implementation of Technology Enabled Affective Learning Using Fusion of Bio-physical and Facial Expression</td>
<td>112–125</td>
</tr>
<tr>
<td>Arindam Ray and Amlan Chakrabarti</td>
<td></td>
</tr>
<tr>
<td>Exploring Teaching Programming Online Through Web Conferencing System: The Lens of Activity Theory</td>
<td>126–139</td>
</tr>
<tr>
<td>Ünal Çakıroğlu, Mehmet Kokoç, Elvan Kol and Ebru Turan</td>
<td></td>
</tr>
<tr>
<td>Usefulness of Social Network Sites for Adolescents’ Development of Online Career Skills</td>
<td>140–150</td>
</tr>
<tr>
<td>Mariëlle Ruten, Anje Ros, Marinka Kuijpers and Karel Kreijns</td>
<td></td>
</tr>
<tr>
<td>A Mobile Application to Improve Learning Performance of Dyslexic Children with Writing Difficulties</td>
<td>151–166</td>
</tr>
<tr>
<td>Rabbia Tariq and Seemab Latif</td>
<td></td>
</tr>
<tr>
<td>Promoting Student Metacognition through the Analysis of Their Own Debates. Is it Better with Text or with Graphics?</td>
<td>167–177</td>
</tr>
<tr>
<td>Marc Lafuente Martínez and Ibis M. Álvarez Valdivia</td>
<td></td>
</tr>
<tr>
<td>Advancing Adventure Education Using Digital Motion-Sensing Games</td>
<td>178–189</td>
</tr>
<tr>
<td>Ju-Ling Shih and Yu-Jen Hsu</td>
<td></td>
</tr>
<tr>
<td>Evaluation of Intercultural Instructional Multimedia Material on Implicit Xenophobic Cognition: Short Time Effects on Implicit Information Processing</td>
<td>190–202</td>
</tr>
<tr>
<td>Joerg Zumbach, Gerhard Schragl, Chad Mortensen and Stephanie Moser</td>
<td></td>
</tr>
<tr>
<td>The Effect of the Digital Classroom on Academic Success and Online Technologies Self-Efficacy</td>
<td>203–212</td>
</tr>
<tr>
<td>Mehmet Arif Özerbas and Bilge Has Erdogan</td>
<td></td>
</tr>
<tr>
<td>The Effectiveness Evaluation among Different Player-Matching Mechanisms in a Multi-Player Quiz Game</td>
<td>213–224</td>
</tr>
<tr>
<td>Fu-Hsing Tsai</td>
<td></td>
</tr>
</tbody>
</table>
**ABSTRACT**

The purpose of the present study was to investigate the effect of captioned vs. non-captioned instructional videos on the motivation and achievement. To this end, a pre-test and post-test experimental design was used on 109 sophomores from a Turkish state university. Videos with and without captions of the unit in question were prepared by the lecturer of the course “Graphics and Animation in Education.” The first group (captioned video group; \( n = 57 \)) studied the applied dimension of the course with captioned videos but the second group (non-captioned video group; \( n = 52 \)) studied without captions. Qualitative and quantitative data were collected for study. Qualitative data were collected via academic achievement test and instructional materials motivation survey while quantitative data were collected by means of focus group interviews to substantiate the quantitative data. The findings indicated that, in contrast to the suggestion of the redundancy principle, motivation and achievement scores of students do not vary according to the instructional video type under investigation (captioned vs. non-captioned). Thereby, it was concluded that a moderating effect of the streaming feature of instructional material should be considered to interpret the redundancy effect. However, further research is needed to better reveal this moderating effect.

**Keywords**
Redundancy principle, Multimedia, Captioned videos, Motivation

**Introduction**

As advances in educational technology can be of considerable benefit, they may inhibit rather than facilitate learning unless they are adapted to the human cognitive system (Leslie, Low, Jin, & Sweller, 2012). To achieve such an adaptation, it is required in educational settings to convey instructional messages to students effectively. The instructional messages can be communicated audibly and/or visually to the students (Kalyuga, 2012) by means of interactive multimedia that presents instructional materials in various combinations of on-screen texts, images, video, audio, and animation (Adesope & Nesbit, 2012). Many instructional designers frequently suggest that both audible and visual presentation of the same verbal information is likely to enhance learning. However, theoretically speaking, it is claimed that if students are provided with spoken and written texts simultaneously, the working memory will be overloaded (Kalyuga, Chandler, & Sweller, 2004; Sweller, Merrienboer, & Paas, 1998). This is called the “redundancy effect” in the related literature. The majority of studies on the redundancy effect used spoken-written presentations (i.e., audiobooks, lectures added spoken-written narrations by instructors, as well as subtitled television programs/videos) (Adesope & Nesbit, 2012).

Multimedia spoken-written presentations can be classified according to whether they have streaming or not. This streaming feature provides more interactivity to presented content compared to other educational presentations (Hartsell & Yuen, 2006). Thus, they help in understanding complex concepts and procedures that cannot be explained in a simple way with text and graphics (Klass, 2003). Instructional videos and TV programs can be given as examples of full streaming formats in which educational content is entirely presented to students as full streams. Other presentation formats (i.e., animation and audiobooks but not videos and TV programs) can be given as examples of non-streaming or partial streaming formats. The partial streaming format contains at least one video clip.

In the literature on the redundancy effect, few studies have investigated full streaming presentations such as instructional videos (e.g., Adesope, 2010; Jadin, Gruber, & Batinic, 2009; Linebarger, 2001; Perez, Van Den Noortgate, & Desmet, 2013; Yüksel & Tanrıverdi, 2009; Winke, Gass, & Sydorenko, 2010). However, there are many studies that reported on educational non-streaming or partial streaming presentations such as animation and concurrent narration, spoken narration with the printed text and documents consisting of diagrams and spoken information (e.g., Mousavi, Low, & Sweller, 1995; Mayer, Heiser, & Lonn, 2001; Moreno & Mayer, 2002; Craig, Gholson, & Driscoll, 2002; Craig, Driscoll, & Gholson, 2004; Kalyuga et al., 2004; Chu, 2006;
Instructional videos

Instructional videos can be described as powerful media that can provide narrative visualization and engage multiple senses of the students simultaneously (Palmer, 2007). The “watch again” feature of instructional videos provides a variety of control elements such as “play,” “pause” and “stop” is an important step in learning, which can help learners enrich their notes on the subject (Whatley & Ahmad, 2007). Instructional videos also not only provide an environment for the learners to learn at the speed, time and setting they can decide, but also allow them to learn independently of others (Chan, 2010). Instructional videos, the preparation and dissemination of which has become quite easy thanks to advancements in modern information and communication technologies, play an important role in auditory and visual transfer of the instructional message to the learner. Videos are now widely used for the dissemination of the instructional content of applied sciences and language teaching (e.g., http://englishcentral.com; Aldera & Mohsen, 2013; Dufour, Cuggia, Soula, Spector, & Kohler, 2007). For example, the efficacy of instructional videos in physics, chemistry, biology, engineering, medicine and anatomy has been the focus of many researchers (e.g., Constantinou & Papadouris, 2004; Palmer, 2007; Aronson, Plass, & Bania, 2012; Hakkarainen, 2009; El-Sayed & El-Raouf, 2013; Holland, Smith, McCrossan, Adamson, Watt, & Penny, 2013). Similarly, instructional videos have enriched the curricula of such applied fields as teacher training (e.g., Moreno & Ortega-Layne, 2008; Borko, Jacobs, Eiteljorg, & Pittman, 2008; So, Pow, & Hung, 2009), software programming, and graphic and web design.

Using captions in instructional videos

Captions are word-for-word translations of dialogues in videos, and can generally be inserted in videos in consideration of the figure-ground relation (Udo & Fels, 2010). Audiences mostly claim that captions are quite motivating in understanding the narrated and displayed content (Danan, 2001; Koolstra & Beentjes, 1999; Koskinen, Wilson, Gambrell, & Neuman, 1993). Nowadays, captions can be added easily to instructional videos by instructors or instructional designers by means of programs like Adobe Premier and Camtasia. Captions in instructional videos provide multisensory information allowing students to view the narration and written words in a synchronized way (Aldera & Mohsen, 2013). Due to all these aforementioned benefits, the use of annotations and captions in instructional videos to facilitate comprehension has been adopted by many teachers and researchers (Yüksel & Tanrıverdi, 2009).

Redundancy effect in captioning instructional videos

According to the cognitive theory of multimedia learning, there are different processing systems for visual and verbal information, and learning occurs by establishing connections between visual and verbal channels each of which has a limited processing capacity (Jamet & Le Bohec, 2007). The main claim of the redundancy principle, which is based on the cognitive theory of multimedia learning, is that students learn better when animation/illustration and narration are presented without written texts than when animation/illustration and narration are presented together with written words (Mayer, 2001; Mayer, 2005; Mayer, 2009; Mayer & Moreno, 2002; Mayer & Moreno, 2003). In other words, the redundancy effect suggests that the animated/illustrated and narrated presentation of an instructional video content affects the learning process more positively in contrast to an animated/illustrated and narrated presentation accompanied by captions or other written texts.

Most studies performed on captioning instructional videos did not support the redundancy principle. For example, in a study conducted by Linebarger (2001), it was found that audio and captions used in videos helped students to focus on the central elements by triggering their reading efforts. A study that utilized instructional material prepared by means of Microsoft PowerPoint, and then converted to instructional video format, revealed that students in the animation + narration + caption group got higher scores in physics achievement tests. They
also displayed an increase in the level of their interest according to the physics inventory compared to those of other treatment conditions (animation + captions only, animation + narration only, conventional lecture method) (Adegoke, 2010). In addition to these studies, Perez et al. (2013) investigated the effects of watching captioned instructional videos on listening comprehension and vocabulary in foreign language teaching, and they found considerable influence of captions in both cases. In another study, Winke et al. (2010) concluded that captioned videos were more effective than non-captioned videos in foreign language listening activities. In a study conducted on the use of video-based e-lecture presentations, Jadin et al. (2009) discovered that the presentation mode of the learning content (presentation of content with vs. without synchronized written transcript of the oral presentation) did not affect students’ learning outcomes. Similarly, Yüksel and Tanrıverdi (2009) did not find a significant difference between the achievement levels of student groups after watching a movie with vs. without captioning.

As shown in the literature, when the achievement levels of the captioned and non-captioned video groups were compared, either the first group was more successful than the other, or there was no meaningful difference between them. Therefore, it would not be wrong to say that more studies are needed in reference to the redundancy effect on instructional videos. From this viewpoint, the current study aims to investigate, in consideration of the redundancy principle, the effect of using captioned instructional videos as supplementary course material on the academic achievement and motivation of university students. To this end, this study sought answers to the following research question: “Do the motivation and achievement scores of students vary according to the captioned or non-captioned instructional videos?”

Method

Design and participants

The experimental design was used in the study. In this respect, the dependent variables were academic achievement and motivation, while the independent variable was video-assisted instruction (captioned vs. non-captioned). The participants were 109 sophomores of the Computer Education and Instructional Technology Department at Çanakkale Onsekiz Mart University in Turkey. Fifty-seven students (34 male, 23 female, average age = 21) comprised the caption group, while the non-caption group consisted of 52 students (34 male, 18 female, average age = 21). Participants did not know whether they were assigned to a “caption” and “non-caption” group. All of the participants had computers and internet access at their place of residence (flats, dormitories, etc.). At the end of the intervention, an achievement test and a motivation survey were administered to both groups. Additionally, qualitative data were collected to substantiate the quantitative findings. More specifically, a focus group interview was conducted with the participants. Eight voluntary students (4 students from each group) participated in the focus group interview, which was voice-recorded.

Data collection tools and materials

Achievement test (applied test)

A seven-item applied test was developed for evaluation of the unit “The Use of Adobe Flash CS5 Tools and Basic Animation Methods.” Cronbach’s Alpha internal consistency coefficient was found to be $\alpha = .76$ for the pretest and $\alpha = .71$ for the post test. Two lecturers carried out the marking independent of each other.

**Soru 4 (15 Puan)**

Yandaki resme tıklayınız ve animasyonu izleyiniz.

Bu animasyonu Flash Programında Shape Tween Animasyon olayını kullanarak hazırlayın.

*Figure 1.* A sample question from the applied test [Question 4 (15 points)/Click on the following image and create this animation by using Shape Tweening on Flash Program]

As a result of this process, inter-rater reliability of the raters was found to be .95. According to Nunnaly and Bernstein (1994), a value over .70 indicates that the internal consistency of the scale has been achieved. The
Motivation and learning are two concepts that complete each other, regardless of the environment (Schunk, 1996). Besides, the design of a given instructional environment may affect both motivation and learning (Dennen & Myers, 2010). From this perspective, each component of the ARCS model, which offers a systematic model for motivating instructional design (Orey, McClendon, & Branch, 2006), provides principles and strategies to increase learning motivation (Dennen & Myers, 2010). The ARCS model has four basic components consisting of the following (Keller, 1983). The [A]ttention component includes strategies to maintain and improve students’ curiosity and interest. The [R]elevance component involves strategies associated with students’ needs, interests and motivations. The[C]onfidence component comprises strategies that will ensure the continuous attendance and achievement of students. Finally, the [S]atisfaction component contains strategies that offer students internal and external reinforcements.

The Turkish version (Kutu & Sözbilir, 2011) of the ARCS Instructional Materials Motivation Survey originally developed by Keller (1987) was used to evaluate the effect of the use of supplementary captioned videos for the course on students’ motivation. While the original survey contained four dimensions, namely attention, relevance, confidence and satisfaction, the Turkish adaptation covered only two dimensions (attention-relevance and confidence-satisfaction) with a Cronbach’s Alpha value of $\alpha = .83$. An assessment tool might be reliable in a particular setting, yet unreliable in another setting or when used for another purpose (Vockell & Asher, 1995). Therefore, a priori application was carried out to explore whether the Turkish version would yield reliable results in the target group. In this a priori application, the ARCS motivation survey was administered on 111 juniors who were studying in the same department as the target group at the time of data collection. The a priori application yielded a Cronbach’s Alpha internal consistency coefficient of $\alpha = .90$ (.86 for attention-relevance and .87 for confidence-satisfaction), which signified that the survey was reliable (Nunnaly & Bernstein, 1994), and could be administered on the target group. In the current study, Cronbach’s Alpha internal consistency coefficient was calculated to be $\alpha = .92$ (.85 for attention-relevance and .88 for confidence-satisfaction). Two sample items of the Instructional Materials Motivation Survey were “I could not really understand quite a bit of the material in this lesson.” and “The way the information is arranged on the pages helped keep my attention.”

Focus group interview form

The focus group interview form consisted of four open-ended questions. The form contained items about “the effect of the presence or absence of captions in the instructional video on the participants’ motivation and learning” and “which one they would prefer, and why.”

Captioned and non-captioned videos

Applications in the unit of “Using Adobe Flash CS5 Tools and Basic Animation Methods” course were video-recorded by the lecturer to be used in the course “Graphics and Animation in Education.” The videos were captioned by the same lecturer. The videos were captioned for the captioned video group but not for the non-captioned one (Figure 2). Separate video segments for each application of the unit were packaged to create learning objects in SCORM standards, and uploaded to the Moodle Learning Management System to collect their tracking data.

Implementation process

The implementation process took a total of six weeks. In the first week, an applied test (pretest) was administered both to the captioned video group and the non-captioned one to assess their initial achievements in the unit. No significant difference was found between the achievement scores of the groups ($t(107) = -1.43, p < .05$), which meant that they did not have prior knowledge. From the second week on, the instructor lectured...
about the theoretical dimension of the unit for four weeks. The lecturer worked face-to-face with both groups in the first two hours of the weekly four-hour course. During the remaining two hours, the students studied only with instructional videos on the Moodle Learning Management System. In the sixth week following the completion of the unit, firstly a motivation survey and then an achievement test were administered to the students in the computer laboratory. A focus group interview was also conducted on the participants at the venue and mutually agreed upon by them and the researcher. The interview lasted 1 hour and 20 minutes.

Data analysis

SPSS was used for the analyses of the quantitative data, and NVIVO for the analyses of qualitative data. One way MANOVA was used to find out whether the motivation and achievement scores of students varied with the captioning or non-captioning conditions of the instructional videos. Qualitative data were analyzed via a descriptive analysis method because the conceptual structure and themes (effects of captioned and non-captioned videos on academic achievement and motivating factors) of the study were determined in advance. The simultaneous use of quantitative and qualitative data increases the validity and reliability of a study (Creswell, 2013). The qualitative data were analyzed by two independent researchers. It was concluded that reliability (Miles & Huberman, 1994) was 84% for the first question, 93% for the second, 91% for the third and 86% for the fourth question. Quotes were given in the following format (Captioned/Non-captioned group, Participant [1-8], temporal data).

Findings

The mean score of academic achievement was 62.18 ($SD = 21.12$) for the captioned group whereas it was 61.66 ($SD = 24.24$) for the non-captioned group. On the one hand, the mean score of motivation was 4.13 ($SD = .55$) for the captioned group while it was 3.92 ($SD = .49$) for the non-captioned group. Given the motivation dimensions, attention-relevance dimension’s mean score was 4.25 ($SD = .53$) for the captioned group and 4.11 ($SD = .56$) for the non-captioned group. In addition, confidence-satisfaction dimension’s mean score was 4.02 ($SD = .63$) for the captioned group and 3.77 ($SD = .53$) for the non-captioned group.

The one-way MANOVA revealed that the motivation level and achievement scores of the participants did not vary with or without the use of captions in instructional videos. ($F(2, 106) = 2.07$, $p = .132$; Wilks’ Lambda = .962; partial eta squared = .038; observed power = .417). In other words, the motivation and achievement scores of the caption and non-caption groups showed no significant difference.

According to qualitative findings about achievement, the participants summarized the probable effects of the captioned and non-captioned instructional videos on their learning process as follows; "We may not have understood some parts because the videos didn’t have captions. Then, we had to rewind the video and watch the unintelligible parts over and over. We listened to [the narration] again. In the end, we could understand but
According to the qualitative findings of the study, the effect of the captioned and non-captioned instructional videos on the participants’ motivation level was as indicated in Figure 3.

Figure 3. Effects of use of captions in instructional videos on learners’ motivation

Some students remarked that captions facilitated the understanding of a hard-to-understand segment. One of them said, “I was in the non-caption group, but it would be better if the video were captioned... The teacher streamed one part of the video so fast that I probably watched that part over and over for about 5 minutes. If it was captioned, then I would have rewound just once, and I would have done it by following the captions and successfully completed the application.” (Non-captioned group, Participant 8, 17:30).

Some students (n = 2) warned that captions might cause distractions in some cases, while some others (n = 5) stated that captions might play a positive role on sustaining attention in lectures. One such student said about the effect of captions on enhancing attention level and span that “Because it is an applied subject if you got distracted, you would not be able to come back. Just relying on the sound might make it difficult to follow the course, even if the teacher showed us on the screen. Captions would facilitate keeping up with the flow of the course, and increasing attention level.” (Non-captioned group, Participant 5, 37:00). Another student disagreed with this view suggesting, “…it is true that sometimes going back to unknown words adversely affected us, but I think a wholly captioned video would distract me.” (Non-captioned group, Participant 7, 29:58).

On the motivating effect of captions regarding prevention of misunderstanding, one participant offered the following opinion: “Sometimes it might be hard to understand whether the teacher said “Ctrl + D” or “Ctrl+ B”. We could clearly understand what he said when we checked out the captions. This is pretty comforting and motivating...” (Non-captioned group, Participant 8, 31:30), says he.

Even though the number of those who stated that captions motivated them was quite high, there were also some complaints about the adverse effects of captioning. Some suggested that students should be able to make the captions visible or invisible as they desired. One said, “…it is rather difficult to follow the instructions while doing what is said. It would be better if the captions could be hidable.” (Captioned Group, Participant 2, 58:04).

Discussion and conclusions

This study investigated whether the captions used in instructional videos had a significant impact on students’ achievement and motivation. In contrast to the suggestion of the redundancy principle, it was shown that the motivation and achievement scores of the participants were not significantly influenced by the captioning or non-captioning of instructional videos used in a course. Jadin et al. (2009) (video-based e-lecture) and Yüksel and Tanriverdi (2009) (captioning movie), who used instructional materials with a streaming feature in their studies...
did not find any significant difference in terms of achievement level. In this respect, the findings of the present study supported the findings of these prior studies. In addition, a reverse redundancy effect (Moreno & Mayer, 2002) was revealed in some other studies (e.g., Adegoke, 2010; Linebarger, 2001; Perez et al., 2013; Winke et al., 2010) in which instructional videos with the streaming feature were used. Consequently, the majority of the studies that were conducted on captioning instructional videos did not support the redundancy principle claims.

The research findings of this study could have been affected by the streaming feature of the instructional material used. In their meta-analysis study on verbal redundancy, Adesope and Nespit (2012) found that moderating variables such as the pacing of presentation, degree of correspondence between audio and text, segment size and inclusion of images or animation should be taken into consideration while interpreting the redundancy effect. However, one moderating variable they missed was the streaming feature of instructional material: animation and audiobooks can be examples of non/partial streaming formats while video and TV programs can be examples of full streaming formats. The moderating effect of the streaming feature of instructional material on redundancy should be further studied. From this viewpoint, the present study is considered to be beneficial in bringing about a new perspective on the redundancy principle as well as encouraging further research into the issue under discussion.

The research findings of this study could be affected by the subject type (theoretical or applied). In this study, the redundancy principle was investigated in the context of the instruction of an applied subject (practice regarding use of the Adobe Flash Program). According to Schwartzman and Henry (2009), the two elements that underlie applied learning are “concrete experience” and “learning by doing.” However, a review of the literature showed that almost all previous studies focused on the teaching of theoretical subjects (e.g., Chu, 2006; Craig et al., 2002, Experiment 2; Diao et al., 2007; Diao & Sweller, 2007; Gerjets et al., 2009; Jamet & Le Bohec, 2007; Kalyuga et al., 2004, Experiment 3; Mayer et al., 2001, Experiment 1; Rias & Zaman, 2010; Adegoke, 2010; Chang et al., 2011; Debuse et al., 2009; Jadin et al., 2009; Linebarger, 2001; McNell, 2004; Montali & Lewandowski, 1996; Perez et al., 2013; Samur, 2012; Winke et al., 2010; Yüksel & Tanrıverdi, 2009). Therefore, research investigating the redundancy principle in the context of subject type (theoretical or applied) is a necessity.

According to the qualitative data, it was found that captions help motivate students by enhancing the comprehensibility of the narration and keeping the students alert. Therefore, it can be suggested that the unintelligibility of verbal statements in non-captioned videos cause students to feel somewhat insecure. On the contrary, students feel more motivated if captioned instructional videos are used in a course because they can more easily catch up with the flow of the narration. Findings that reveal that captions are both distracting and comforting at the same time can be perceived as contradictory, as one student’s suggestion that “the captions should be visible if need be” revealed. The qualitative findings regarding motivation do not support the quantitative findings (No significant difference was found.). Although motivation score in the captioned group \( (M = 4.13; SD = .55) \) was higher than that in the non-captioned group \( (M = 3.92; SD = .49) \), the difference was not significant. In addition, the value pertaining to the statistical power of the finding is very low (power < .80). From this perspective, further research is needed to investigate the motivation and achievement variables in similar studies.

Mayer and Johnson (2008) stated that the studies on the redundancy principle were carried out as short-term laboratory experiments and that this issue should be investigated in more realistic educational settings. Accordingly, this study was conducted in a realistic educational setting and, in a longer term, which differentiates it from other similar studies in the literature. Furthermore, the instructional video used in this study was intrinsically learner-paced. To put it more clearly, the theoretical parts of the subjects were presented face-to-face and the application sections were taught through instructional videos.

The participants’ repeated watching of the videos might have contributed to their learning process. Also, face-to-face lecturing of the theoretical dimension of the subject and/or students’ re-watching the videos might have eliminated the probable adverse effects of the captioned videos. Therefore, the data collected in this study on the effects of captioned and non-captioned videos on learner achievement and motivation needs to be substantiated by further research, which can additionally investigate video narration of theoretical dimensions.

**References**


Computer Card Games in Computer Science Education: A 10-Year Review

Maria Kordaki* and Anthi Gousiou

Department of Cultural Technology and Communication, University of the Aegean, University Hill, Mytilene, Greece // m.kordaki@aegean.gr // goussiou@aegean.gr

*Corresponding author

(Submitted March 26, 2015; Revised November 28, 2015; Accepted January 4, 2016)

ABSTRACT

This paper presents a 10-year review study that focuses on the investigation of the use of computer card games (CCGs) as learning tools in Computer Science (CS) Education. Specific search terms keyed into 10 large scientific electronic databases identified 24 papers referring to the use of CCGs for the learning of CS matters during the last decade. The findings revealed that the CCGs reported by the reviewed papers were used for the learning of diverse CS issues. The motivation behind the use of the aforementioned CCGs was based on: (a) students’ familiarity with the CCGs at hand, so that they might focus on the construction of their simulation, (b) the fact that, the logic of CCGs is suitable for the teaching of various CS concepts, and (c) CCG-play as a motivational activity to engage learners in their learning. CCGs were used in two ways: as CCG-construction context based on supportive data provided by educators, and as CCG-play educational tools. The constructivist learning approach was implied behind the design and use of most of the CCGs reported in the articles reviewed. Evaluation of some CCGs also provided positive results.

Keywords

Computer card games, Computer science, Review, Education

Introduction

Computer games, a significant part of human social and cultural environment, seem to attract people’s interest, attention and energy (Oblinger, 2004) by enabling players’ participation in new, otherwise, inaccessible worlds (Shaffer, Squire, Halverson, & Gee, 2005). Furthermore, players’ engagement is enhanced by essential game characteristics, namely, interest, fun, challenge, fantasy, curiosity, and control (Malone, & Lepper, 1987), hence, “flow” experience could be enabled (Csikszentmihalyi, 1990). Researchers have acknowledged that Educational Computer Games (ECGs) have the potential to be effective tools for the learning of challenging and complex matters because they are able to: (a) create personal learner motivation and enjoyment, (b) support multiple learning styles and skills, (c) enable active learning, (d) reinforce mastery skills, (e) provide an interactive, problem solving and decision making context, and (f) support exploration and experimentation in a protected environment and learning from the results (Kebritchi, & Hirumi, 2008; Oblinger, 2004). Researchers have also suggested that apart from the game-play learning approach students could construct their own computer games in order to explore new learning concepts in active, meaningful, engaging and effective ways, consider the digital culture from the producer’s perspective instead of that of the consumer/player (Kafai, 2001; Ke, 2014) at the same time developing computing and problem-solving skills (van Eck, 2006).

Computer Card Games (CCGs) are the modern expression of Card games (CGs) that is a genre of games which have been around for centuries. Specifically, historians have traced playing cards back to the 10th century and most game taxonomies include the large class of CGs (Crawford, 1982). CGs have been also used for educational purposes since the time of Piaget (Kamii & DeVries, 1980). Furthermore, CGs has been used to help students understand and apply basic aspects of developmental psychology and especially the stages of Piaget’s theory of cognitive development (Weisskirch, 2003). CGs are also seen as simple games, where players use cards with specific characteristics and concentrate on the forming of appropriate card combinations (Crawford, 1982) while taking into account probability, classification, grouping, comparison and matching issues. In fact, classification activities are central to CG play since players understand the learning concepts under discussion by forming appropriate groups of cards, thus promoting critical thinking, reflection and problem-based learning (Kordaki, 2015). In addition, due to their simplicity in equipment and rules, educational CGs could be used as the entry point for motivating and engaging novice players in the game-based learning experience (Gosper, & McNeil, 2012). Motivation and learning effectiveness are also supported by the competitive elements of CGs. In addition, well-structured CGs can integrate instructional content with the game rules so as to make abstract and complicated concepts intuitive and unambiguous for learners (Su, Cheng, & Lin, 2014). Moreover, appropriately-designed CGs have the potential to improve learners’ communicative skills and to promote active learning through interaction with other players (Bochennnek, Wittekind, Zimmermann, & Klingebiel, 2007). At
the same time, CGs could also be used to enhance the enforcement of memorization, matching, number manipulation and pattern recognition skills (Oblinger, 2004) while they might also encourage learners’ logico-mathematical and interpersonal intelligence (Berger, & Pollman, 1996). As far as educational computer card games (ECCGs) are concerned, in several cases further perspectives on learning and e-learning have emerged through the adoption of new modalities of interaction during CG play. Another noteworthy feature that seems to contribute to the knowledge construction of learners concerns the innovative pedagogical approaches taken into account in the design of several ECCGs (Kordaki, & Gousiou, 2014). Constructivist methodologies have also been proposed for the design of ECCGs that acknowledge problem solving and students’ misconceptions (Kordaki, 2015). CCGs have been used to support the learning of various subjects, such as: mathematics, science, language, history, environmental education, health (Kordaki, 2011; Gousiou, & Kordaki, 2015).

As far as CS education is concerned, several ECGs have been suggested to support the learning of various CS subjects, such as programming, binary system, and hardware issues (Kazimoglu, Kiernan, Bacon, & MacKinnon, 2012; Kordaki, 2011; Papastergiou, 2009), while a number of projects have been organized to support CS education (Zhang, Kaufman, & Fraser, 2014). Computer Games have been used in many different ways in CS education (Hakulinen, 2011) including games as: (a) motivators for the learning of CS topics, (b) artefacts to be constructed by the students; since computer game development encompasses many aspects of CS, e.g., computer graphics, artificial intelligence, human-computer interaction, security, distributed programming, simulation, and software engineering (Overmars, 2004), game-design has been suggested as a vehicle to teach CS issues, (c) examples to teach CS topics, and (d) learning environments to learn CS issues through game-play (Wallace, McCartney, & Russel, 2010; Zhang, Kaufman, & Fraser, 2014). Despite the above, research into the use of CCGs in CS education has not yet been reported in the literature.

Here it is worth noting that, although it has been acknowledged that some games can foster positive educational outcomes for a variety of learners in specific subjects, this conclusion cannot be generalized to all games in all learning areas for all learners (Hays, 2005). In addition, despite the fact that, the motivational effectiveness of computer games has been supported by the findings of various empirical studies (Facer, 2003), the educational effectiveness of computer games aiming at concrete learning objectives for specific learning subject curricula is still under-researched (Papastergiou 2009). Furthermore, in some cases the empirical results seem to be in contradiction (Kafai 2001; Kirriemuir, & McFarlane 2004) with computer game-based learning providing equivalent results to traditional face-to-face teaching (e.g., Randel, Morris Wetzel, & Whitehill, 1992). Furthermore, concern has been raised over the drawbacks of learning through play, especially if learning becomes “too much fun” (DeVries, 2004) while key questions have been raised about the role of play in the learning of specific subject curricula (McFarlane, Sparrowhawk, & Heald, 2002). Thus, it is essential to investigate the impact of specific types of games – such as CCGs – in specific learning areas, such as CS education. Consequently, in this paper an endeavour has been made to investigate the use of CCGs as learning tools in CS education by reviewing the literature of the last decade. Such a review has not yet been reported.

In what follows, the methodology of the review is introduced, along with the presentation of results. Finally, there is a discussion of the results and conclusions are drawn.

**Methodology**

**Research question**

The central question that this review paper attempts to answer is whether and how CCGs were used as learning tools in CS education. Thus, the objective of the paper is to review the literature of the last decade on the use of CCGs in CS education with the intention to:

- present an overview on the educational context and the ways CCGs are used in CS education
- identify essential CCGs’ specifics
- identify the potential benefits of CCGs in CS education,
- present a synthesis of the empirical evidence available thus, far on the educational effectiveness of CCGs in CS education
- explore the pedagogical framework used based on the overviewed literature

12
Data collection

Databases searched

Ten large electronic databases, which are identified as relevant to education, digital technology and social science were searched in this review, namely: ACM, EdITLib, ERIC, IEEE, Mary Ann Liebert, MIT Press, Oxford University Press (Journals), ScienceDirect, SpringerLink, and Wiley.

Search terms

The search was conducted using the keywords “card game” AND “education” AND “computer.” The search was limited to date from January 2003 to December 2013. As a result, 1,297 papers were identified.

Selection of papers for inclusion in the review

A number of further criteria were specified to select appropriate studies for inclusion in the review. To be included in the review, papers had to: (a) be related to CCGs used in education: After screening the previously identified papers by title and abstract and excluding those ones which were not referred to CCGs used/design for educational purposes, 53 papers found as relevant to ECCGs, (b) not related to the same ECCG and not reporting the same results (2 papers were excluded), (c) related to CCGs used in CS Education, and (d) have the typical form of a scientific paper (including an abstract, theoretical background, game or project overview/architecture/implementation, evaluation/research methodology (in case an evaluation was conducted), results/conclusions/summary and references). Using these four conditions, 24 papers met the inclusion criteria and were identified as relevant for full text review.

Data analysis

Coding of papers

The 24 papers meeting the inclusion criteria were coded using a data extraction proforma that was developed by considering the aforementioned research questions as well as previous research (e.g., Mikropoulos, & Natsis, 2011), which categorized educational digital artefacts, their outcomes and impacts along several salient dimensions. Thus, the said ECCGs were categorized according to ECCGs’:

(a) educational context they dedicated for use, namely:
- Purpose of the game: ECCGs were coded according to whether the game was originally designed as a game for learning through game-play or as a game for learning through game-construction.
- Educational level: ECCGs were categorized according to whether the game was originally designed as a game for Primary, Secondary, Tertiary education and general users.

(b) specifics:
- Subject Discipline: ECCGs were categorized with respect to the specific knowledge domain/curricular area of CS that the game addressed. A brief description of each game was also attempted.
- Motivation behind ECCGs selection as a learning tool.
- Platform/delivery: The platform for delivery of the ECCGs game was categorized as, PC, online game, and mobile.
- Programming language used.

(c) impact on learning
- Evaluation: mentioned/not mentioned.
- Method of evaluation: qualitative, quantitative, and mixed.
- Sample used.
- Outcomes: learning outcomes, students’ attitudes towards ECCGs.

(d) pedagogical context, namely:
- Learning theory explicitly reported, e.g., Constructivism, behaviourism.
- Learning theory implicitly reported: active learning, learning-by-doing, experiential learning, learning-by-example, problem-based learning, collaboration, scaffolding, and reflection.
The papers were individually analyzed by the two researchers/authors of the paper using the previously mentioned coding system. When a disagreement about the classification of papers appeared the researchers collaborated in order to arrive in a mutual agreement.

Results

Table 1 provides a summary of the overviewed research articles in terms of the following attributes: (a) column 1: Articles’ reference; (b) column 2: Educational Level of ECCG (Pr = Primary, M = Middle, T = Tertiary, P = Public); (c) column 3: ECCG’s subject related to CS; (d) column 4: ECCG’s Purpose, Game-Play (G-P) or Game-Construction (G-C); (e) column 5: ECCG’s Pedagogical framework (C = Constructivism, B = Behaviourism, R = Reflection, AL = Active Learning, LE = Learning-by-Example, LD = Learning-by-doing, PL = Problem-based Learning, CL = Collaboration, EL = Experiential Learning, S = Scaffolding, and (f) column 6: whether or not an evaluation study was conducted.

Table 1. Overview of the reviewed articles

<table>
<thead>
<tr>
<th>Reviewed articles</th>
<th>ECCGs</th>
<th>Education level</th>
<th>Learning subject</th>
<th>Purpose</th>
<th>Pedagogical framework</th>
<th>Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kordaki (2011)</td>
<td>Pr</td>
<td>Binary System</td>
<td>G-P</td>
<td>C, AL, S, R</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Mitsuhara et al. (2007)</td>
<td>Pr</td>
<td>Binary System</td>
<td>G-P</td>
<td>B</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Rodger et al. (2012)</td>
<td>M</td>
<td>Interdisciplinary: Maths &amp; CS</td>
<td>G-C, G-P</td>
<td>C, LD, R</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Grissom, &amp; Dulimarta (2004)</td>
<td>T</td>
<td>OOP</td>
<td>G-C</td>
<td>C, LD, CL, R</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Popyack (2009)</td>
<td>T</td>
<td>AI</td>
<td>G-C</td>
<td>C, LD, PL</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Hunt, &amp; Willison (2011)</td>
<td>T</td>
<td>Concurrent Programming</td>
<td>G-C</td>
<td>C, LE, LD</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Premkumar, &amp; Coupal, 2009</td>
<td>T</td>
<td>Software Development</td>
<td>G-C</td>
<td>C, LD, CL, R</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Wick et al. (2004)</td>
<td>T</td>
<td>OOP</td>
<td>G-C</td>
<td>LE</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Friese, &amp; Rother (2013)</td>
<td>T</td>
<td>AI</td>
<td>G-C</td>
<td>C, LD, R</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Kiss (2013)</td>
<td>T</td>
<td>Programming</td>
<td>G-C</td>
<td>C, LE, LD, R</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Chaves-González et al. (2007)</td>
<td>T</td>
<td>Genetic Algorithms</td>
<td>G-P</td>
<td>C, R</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Chang, &amp; Sung (2008)</td>
<td>T</td>
<td>Operating Systems</td>
<td>G-P</td>
<td>C, PL, R</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Monsalve et al. (2011)</td>
<td>T</td>
<td>Software Engineering</td>
<td>G-P</td>
<td>PL, R, CL</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Albraikan et al. (2013)</td>
<td>P</td>
<td>Haptic Technology</td>
<td>G-P</td>
<td>EL</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

Educational context

Purpose of ECCG

The reviewed articles suggest that CCGs were designed to be used in two main modes, namely game-construction (16 papers) and game-play (8 papers). Concerning the first mode, CS students had the task of implementing specific CCGs in order to learn and assess computational issues in a meaningful and appealing way. The aforementioned CCGs were usually simulations of known CGs, such as “Blackjack,” “25,” “Set,”
“War” and other solitaire games. Under the second mode fall CCGs which were used as educational tools to support student learning through game-play. Some of these CCGs were also simulations of known physical CGs (e.g., “Memory,” “Foursome,” “Blackjack”), created by adjusting the contents of the card-deck to the CS educational subject in question while at the same time making appropriate modifications to the logic and requirements of the CG.

Educational level of ECCGs

The ECCGs under review were suggested for use at various educational levels and for public use, namely: Tertiary (19 articles), middle (1 article), primary level (2 articles), and general users (1 article) while one article reported an ECCG for both primary and tertiary level (see Table 1). It seems that the aforementioned ECCGs were mainly used to enhance learning at tertiary education level.

ECCGs specifics

Subject discipline

The games reported in the reviewed articles are related to the learning of diverse CS subjects, namely: (i) Programming, (ii) Software Development, (iii) Artificial Intelligence (AI), (iv) Binary System, (v) Operating Systems, (vi) Genetic Algorithms, and (vii) Haptic Technology. In the following sections, a brief description of the ECCGs fall in the aforementioned subject disciplines is presented.

Programming: Most of the reviewed articles (12) report on the use of CGs as a context in which several programming issues, such as general principles, programming language-JAVA, OOP design and philosophy, and concurrency, can be taught, and will be described below.

- **General Programming Principles:** During several programming courses, students were assigned the task of developing the following CCGs: (a) A simulation of the CG “25” and after the program implementation they would also be able to play the game for fun (Parlante, Matuszek, Lehman, Reed, Estell, & Chinn, 2004); (b) A simulation of physical CGs such as “BlackJack,” using php. Students were supported with basic programming elements to implement their CGs and consequently facilitate their acquisition of algorithmic thinking (Kiss, 2013); and (c) the development of a betting algorithm for computers of the “Oriental Casino” CG was suggested, as a meaningful real world project (Park, 2010).

- **Programming Language – JAVA:** Students were assigned to implement a simulation of the “Poker” CG, in JAVA, using an environment developed by their professor as a base material (Santos, 2012).

- **OOP:** During several JAVA programming courses, familiar CGs were selected for the learning process. Students were assigned the task of gradually implementing the following CGs in JAVA, based on given material (reusable objects, part of code, etc.) designed by the professors according to OO principles: (a) The simulation of two physical CGs, “25” and “War,” for the learning of basic OO concepts (Leska, & Rabung, 2004; 2005); (b) Simulations of physical CGs, such as “BlackJack,” “Video Poker” and “War,” so that complicated topics of OOP might be taught, such as inheritance, encapsulation and composition (Grissom, & Dulimarta, 2004); (c) A simulation of “BlackJack,” in order to realize the advantages of OOP (Kouznetsova, 2007); (d) The simulation of the CG “Set,” to introduce advanced concepts of OOP, such as polymorphism and design patterns (Hansen, 2004); (e) Seven design rules – using examples of a “Poker” card abstraction – were also suggested as an appropriate methodology to facilitate the teaching of complicated OOP concepts, such as encapsulation. Examples of appropriate and inappropriate implementation of classes, designed by students and professors, according to the main principles of OO principles, were used (Wick, Stevenson, & Phillips, 2004).

- **Concurrent Programming:** A simulation of a simple CG – “California Speedway” – was used – as a project – to introduce three major issues of concurrency, such as race conditions, deadlock and scheduling, to the students, since this CG introduces the abovementioned issues of concurrency in areas of non CS-based non-abstract context. The said simulation was used during two different courses (Operating Systems and Programming Languages), either as a programming assignment to introduce concurrency or as a standard
example that can be used to compare different types of implementations, such as different programming languages (Hunt, & Willison, 2011; Hunt, & McGregor, 2010).

**Software Development (SD):** The importance of using a specific SD methodology was considered in the following three articles, where specific ECCGs were either suggested as a context of learning through CG-construction or as a learning tool through CG-play. Specifically: (a) A networked, multiplayer CG called “Bug is a lie” was suggested for implementation throughout a SD lab-course, in order for SD methodology – based on the agile SD process Scrum – to be taught. Due to its simplicity, students could focus on the learning process from the beginning and develop the CG on the basis of a proper tooling infrastructure and code skeleton, both provided by the professors (Schroeder, Klarl, Mayer, & Kroiss, 2012); (b) During a project-course in Computer Systems Technology, students were involved in a learning-by-doing SD project for computer-based learning and review purposes for medical and other health professionals, in order to acquire workplace familiarization by experiencing actual problems and attempting to provide solutions. During the said project, simulations of two CGs - a “sorting” and a “solitaire” CCG - were implemented (Premkumar, & Coupal, 2009); and (c) an educational web-based simulation of a combination of a physical board and a CG – called SimuLES-W – was suggested as a teaching tool to support student learning about Software Engineering concepts in a simulation of problems happening in a real project development process. This constitutes a collaborative approach, where students must resort to discussion, putting forward arguments during specific phases of the game-play, in order to approve or reject the cards used (Monsalve, Werneck, & Leite, 2011).

**Artificial Intelligence (AI):** ECCGs that were used as a context in order to enhance the learning of specific AI concepts were reported in two articles: (a) Simulations of “BlackJack” and two simpler versions of it, called “WhiteJack” and “GreyJack,” were used during an advanced AI course, so that main aspects of Markov modelling could be conceptualized by the students, since the uncertainty state that the said CGs integrate and the decisions the players have to make during CG-play were used as a context for the coverage of the learning subject. Students had to implement the optimal strategies related to dealer and player policies using a Markov Decision Process in the form of assignments and, ultimately, use implemented policies in a tournament from a “BlackJack” simulator (Poppyack, 2009); (b) Two CGs were used, “Tecards” and “Rok,” so that concepts of AI might be taught effectively and pleasantly. Students developed the agents of the aforementioned CGs, which could be evaluated in a web-based server implemented by professors (Fries, & Rother, 2013).

**Binary System:** Two ECCGs were suggested for the learning of basic aspects of the binary system by primary level students: (a) An alternative, intelligent version of “BlackJack” was designed. Specific cards featuring binary numbers were designed to support the conversion of binary numbers to decimal ones and vice versa by the students (Kordaki, 2011). (b) A computer-based memory CG was designed, combining RFID technology for the learning of the translation of binary numbers to the corresponding decimal ones. This was considered an appropriate learning method since the 1-1 relationship between binary and decimal numbers can be represented with two playing cards (Mitsuhara, Ogata, Kanenishi, & Yano, 2007).

**Operating Systems:** A web-based ECCG that simulates the operating system processes was designed (Chang, & Sung, 2008).

**Genetic algorithms (GAs):** A CG that demonstrates intelligent behaviour by executing GAs was implemented. This was based on an existing popular CG, called “Ten and a half.” Students had to generate an opponent before the beginning of the game by configuring several parameters related to the GA and see their implementation through the CG. Students could also use the log-files created from the game in order to analyse and improve the genetic behaviour of the game (Chaves-González, Otero-Mateo, Vega-Rodríguez, Sánchez-Pérez, & Gómez-Pulido, 2007).

**Haptic technology:** An edutainment CG game for mobile devices - the “HaptiMemp Game” - was suggested in order to enhance user knowledge and experience of haptic. Its design was based on the traditional “memory” CG, which was enriched with haptic feedback (Albraikan, Badawi, Hamam, & El Saddik, 2013).

**Interdisciplinary/Combinatory approaches:** Interdisciplinary and combinatorial approaches mentioning CS issues were reported in two articles: (a) In order for CS concepts be integrated into the teaching process, two simulations of known CGs, “Set” and “War,” were implemented by teachers using the “Alice” software (Kelleher, 2006). These simulations were constructed to support student learning about several mathematical concepts, such as set theory, combinatorics, randomness, and probability issues. At the same time, teachers were provided with support to experiment with various OOP aspects needed when they used “Alice” for the formation of student projects in their classrooms (Rodger et al., 2012); (b) A simple educational CG, “Quartet,” with
speech-enabled cards, was designed both to support primary level pupil learning of mathematical concepts and also to support the teaching of speech recognition concepts and related problems in the case of university students, as well as to motivate them to develop similar applications. (Nagy, 2006).

Platform/Delivery: All but one CCGs run on desktop systems. In addition, the reviewed CCGs were either standalone (9 papers) or web-based (3 papers) while networked CCGs were also reported in two cases. Moreover, one CCG was stand-alone in the single-player mode and web-based in the multi-player mode while only one CCG run on mobile devices as it is related to haptic feedback. Finally, in 10 papers this kind of information was not mentioned.

Programming language used: Concerning the CCGs that were being constructed by CS students, Java was the most frequently used language, mainly combined with environments that support it (e.g., BlueJ, Eclipse IDE, Swing), while PHP, Visual Studio C#, CSS, HTML, PROLOG, Scheme and MATLAB had also been used in individual implementations. This is rather expected since in almost all of the aforementioned articles the learning subject is Object Oriented Programming. As far as the CCGs, that had been designed in order users to learn by playing, are concerned, they were also implemented by using JAVA and its supporting environments but had been also used other developing tools [e.g., Visual Studio .NET C#, Alice, Android Developer Tool (ADT)] as well as programs for specific needs (e.g., Haptic Effect Preview App – for the development of haptic effects).

Motivation behind ECCGs selection as a learning tool: The ECCGs reported in the reviewed papers have been suggested or selected as learning tools in order: (a) to motivate and engage players/students through a meaningful and appealing way in the learning process of various CS subjects (10 papers), (b) to provide CS students an effective framework for the learning of complicated CS issues, because, AI is necessary in CCG-play (2 papers), basic aspects of OOP are inherent in CCG-construction (e.g., classes and objects, methods, inheritance, encapsulation, composition, design patterns and polymorphism) (5 papers), knowledge about concurrent programming (1 paper) through the construction of simulations of known CGs, knowledge about web-based programming and technologies is necessary to support social interaction that is mainly emphasized during CCG-play, (c) to support problem solving by collecting and arranging the proper cards to solve CS well-defined problems with known solutions, (d) to promote social, rich and constructivist learning experiences, by supporting: active learning (3 papers), the enrichment of the teaching procedure in a collaborative way (2 papers), and through the use of modern technologies, such as haptics and “Alice” (2 papers).

ECCGs’ impact

ECCGs’ evaluation

Method-Sample: Empirical studies had been conducted in 10 of 24 papers, where qualitative (6 papers), quantitative (2 papers) and mixed mode (2 papers) were the research design approaches that were used. Data were mainly collected by questionnaires. Concerning the sample used, participants were university students (5 studies), primary school pupils (2 studies), general users (1 study), and the artefacts of learning, namely, students’ agents, and teachers’ digital games and lesson plans. The size of the sample was mainly small or medium scaled; ranged from: up to 10 (1 study), 10-30 (4 studies); 30-50 (2 studies), 50-100 (2 studies) while in one study it is not mentioned.

Outcomes: Concerning the learning outcomes, four (4) of the ten (10) studies report on comprehension and knowledge acquisition through the CCGs-play or CCG-construction where in all but one (Mitsuhara, Ogata, Kanenishi, & Yano, 2007) of them, learning outcomes seemed to be positive. Additionally, six (6) studies referred to students’ attitudes towards the use of ECCGs in the learning process in terms of approval, enjoyment, motivation, usability. All the aforementioned “motives behind ECCGs selection as learning tools” were also achieved as it was verified by empirical data. However, further research is needed to provide more evidence about the learning effectiveness of CCGs.

Pedagogical context

Learning Theory: Modern social and constructivist theories of learning were explicitly mentioned in the design of just one ECCG (Kordaki, 2011) while in other papers, some terms usually used in these approaches were mentioned (see Table 1, column 5). Behaviourism is also implied in one study. Concerning the articles where
CCGs were the artefact of learning, it seems that they also implied constructivism, since students were asked to learn specific CS concepts by actively solving problems in order to construct specific CCGs.

**Collaboration:** Since collaboration is among the key features of modern social learning theories (Vygotsky, 1974), the reviewed articles were analysed in terms of this principle. In fact, only a few articles reported that students had to collaborate face-to-face and online in order to program and implement specific CCGs assigned by their professors (Grissom, & Dulimarta, 2004; Schroeder, Klarl, Mayer, & Kroiss, 2012). Students also had to collaborate using a common interface in order to learn SE concepts within a web-based collaborative platform that is a combination of a computer board game and a CG (Monsalve, Werneck, & Leite, 2011).

**Reflection:** Reflection is yet another key feature of constructivism (Jonassen, 1994) and consequently special attention has been paid to analysing the reviewed papers according to this metacognitive skill. It emerged that reflection was attained in most cases (17 papers; see Table 1, column 5). In truth, this was attained in two contexts: Firstly, during the construction of CCGs by CS students, where they reflected on their prior knowledge in order to implement these games while at the same time gradually grasping appropriate new CS concepts, and, secondly, during CG-play, where players reflected on their knowledge at the same time as being supported by the CCG at hand as well as by peer-interaction (e.g., Premkumar, & Coupal, 2009; Monsalve, Werneck, & Leite, 2011).

**Discussion and conclusions**

This work explored CCGs that have been used to support the learning of CS concepts during the decade 2003-2013 through a review study based on 24 articles identified by using specific search terms. The vast majority of the articles referred to the use of CCG-construction as learning tool in CS tertiary education, with the exception of 4 papers that reported the use of CCG-play as learning tool in lower educational levels or in public use.

Two different approaches to the way CCGs were integrated into the learning process emerged: CCGs as educational game-play tools where students learn by playing the game at hand, or as the artifact of learning where students had to learn computational issues by designing and constructing either the whole game or a part of it based on supportive data they were given by their educators. These findings are congruent with the four approaches suggested by Wallace, McCartney, and Russel (2010).

However, it would appear that CCGs were not only used to motivate and engage students to learn computational issues as it happens when using other types of computer games. In fact, Professors in CS departments recommend the CCG-construction approach due to the inherent logic of this genre of games. It is claimed and verified by empirical data that this logic is suitable for the meaningful learning of various computational issues such as OOP philosophy and principles (e.g., classes and objects, methods, inheritance, encapsulation, composition, design patterns and polymorphism) as well as networking technologies as social interaction is embedded in CGs-play. It also seemed that a number of CS issues can be learned, through CCG-construction such as: programming, concurrency, database design, SE and AI. It was also emerged that, due to the fact that, CGs are familiar, common and simple games; students, being accustomed to this type of game, can focus on the learning issues without the obstacle of needing to learn the game philosophy. On the other hand, rich, problem solving, social, and constructivist learning experiences have been promoted through CCG-play for the learning of various CS topics such as: binary system, GA, haptic technology, operating systems and software engineering.

The results of the review also show that empirical studies were conducted in less than half of the articles with positive results in terms of learning outcomes and students’ attitudes towards the use of ECCGs in the learning process in terms of approval, enjoyment, motivation, and usability. However, more and large-scaled studies are needed to provide more evidence about the learning effectiveness of CCGs.

Modern constructivist views of learning have been also adopted in the use and construction of the ECCGs reported. These views have been mainly implied by the terminology used in most of the reviewed articles. However, it is worth noting that only one of the reviewed articles clearly stated the learning theory that authors followed in the design of their ECCGs. Reflection was attained in most of the reviewed articles and seemed to support the learning process. However, despite the fact that, social interaction is inherent in CG-play only few studies reported collaboration mainly through CCG-construction.

In conclusion, it seems that the construction of CCGs constitutes a meaningful, appropriate and suitable context that has been mainly used in CS tertiary education for the learning of various computational issues, while only a
few ECCGs have been suggested as educational, game-play tools. This review study also unfolds interesting perspectives for CS educators, game developers, and students needing further research. The adoption of CCG construction in primary and secondary education to support the learning of CS concepts could also attract the attention of CS researchers and educators. In addition, game designers and developers could be prompted to emphasize the design and implementation of ECCGs so that students might learn various CS concepts through CCG-play. By encouraging students to learn CS subjects early in life, which is attainable within the meaningful and compelling context of CCG-play, they would potentially be inspired to choose careers in the field of CS. All in all, it is hoped that these findings will contribute sufficient information to CS educators and game developers to enable them to be effective users or developers of ECCGs.

References


ePortfolio-Based Learning Environments: Recommendations for Effective Scaffolding of Reflective Thinking in Higher Education

Pauline Roberts¹, Dorit Maor² and Jan Herrington³

¹Edith Cowan University, Australia // ²Murdoch University, Australia // pauline.roberts@ecu.edu.au // d.maor@murdoch.edu.au // j.herrington@murdoch.edu.au
³Corresponding author

(Submitted April 1, 2015; Revised January 11, 2016; Accepted January 19, 2016)

ABSTRACT

In addition to providing a useful repository for learning products, ePortfolios provide enhanced opportunities for the development of advanced learning skills. It can be argued, however, that ePortfolios are not being implemented effectively towards fulfilling this important function. This paper presents an investigation of an ePortfolio environment that scaffolded the learning of pre-service teachers. The environment was embedded within the PebblePad platform and utilised the Blog function to provide students with activities that were designed to enhance and support the skills and dispositions required to undertake action research. Prompts were provided to students to scaffold the completion of an action research project and provide additional activities that supported the enhancement of reflective thinking. The research study utilised an eLearning Lifecycle that provided a cyclic framework of review and implementation. The purpose of this model was to identify design principles for future iterations of ePortfolio-based learning environments. Findings suggest that the prompts and the ePortfolio environment were effective in scaffolding students’ reflective thinking. Additionally, design principles are suggested to ensure this research has both practical and theoretical significance for implementation in similar environments.

Keywords

ePortfolio environment, Scaffolding, eLearning lifecycle, Learning environment

Introduction

In recent years, the increased use of electronic portfolios (ePortfolios) has seen the enhancement of the affordances these programs provide. No longer confined to the role of “evidence repository,” ePortfolios now provide features that enable the scaffolded development of advanced learning skills (Barrett, 2006). As part of this development, ePortfolios have moved beyond the traditional purpose of assessment to one that more deeply facilitates and enhances student learning. This maturation of ePortfolio environments, along with recognition of their educational strengths, led to the implementation of this research study that examined the use of the Pebblepad ePortfolio platform as a learning environment.

Portfolios have been used for many years in artistic fields as a means of showcasing skills and abilities through work samples and documentation of performances (Drier, 1997). In education, they have also been used as assessment tools to demonstrate achieved competencies, to allow students to document learning reflections and to understand how they have applied their knowledge (Hartnell-Young et al., 2007; Shepherd & Hannafin, 2011). Portfolios are also strongly aligned with integrated and authentic assessment (cf. Gikandi, Morrow, & Davis, 2011; Herrington, Reeves, & Oliver, 2010).

The development of technology and the increased availability of web-based platforms to document and present work in multi-media formats has led to widespread use of ePortfolios (Clark & Eynon, 2009; Duncan-Pitt & Sutherland, 2006; Khoo, Maor, & Schibeci, 2011). The ePortfolio has been described as, “one of the most adaptable tools currently available to educators around the world” (Littlejohn & Pegler, 2007). One of the key affordances of an ePortfolio, in comparison with its paper-based predecessor, is its ability to contain several layers of evidence that can be used for varying purposes (Lambert & Corrin, 2007) and to portray learning to alternate audiences (Chesney & Marcangelo, 2010). Three important differences between ePortfolios and their paper-based equivalent are:

- Digital formats are easy to rearrange and edit
- ePortfolios allow hyperlinks to connect documents and link to external sources/references
- The electronic format is portable and mobile (Stefani, Mason & Pegler, 2007, p. 17).

Building on these developments, ePortfolio technology has now progressed to a point where it is being suggested as a platform for learning (Barrett, 2005). ePortfolio platforms provide access to tools and resources that can be designed and implemented to guide student learning and allow for collaboration with others in this process.
A great deal of research highlights the potential of the ePortfolio as a learning tool (Housego & Parker, 2009), although the evidence of the effectiveness of these initiatives is still being collected (e.g., Oakley, Pegrum, & Johnston, 2014).

This paper aims to provide this missing evidence by investigating the implementation of an ePortfolio-based learning environment that was used to enhance reflection in pre-service teachers as they completed an action research project. The research was guided by the overarching research question of: How effective is an ePortfolio environment in providing the scaffolding required to enhance reflective thinking in pre-service teachers?

The scaffolding process was initiated within the PebblePad ePortfolio platform. In this context, scaffolding supported the learning process that was tailored to the needs of the students with the aim of helping them achieve their learning goals (Sawyer, 2006). PebblePad was chosen as the ePortfolio platform for this research because: it allowed students to share resources through the Blog function; it used privacy settings so that students’ assets remained private until they chose to share them; and it offered options for sharing and collaboration of assets. The environment was designed from the perspective of an Enculturation Teaching Model (Tishman, Jay, & Perkins, 1993) that utilised a series of prompts to detail activities designed to target reflective skills and dispositions that could assist the completion of an action learning project.

To evaluate the effectiveness of this environmental approach, the research study involved a cyclic implementation and review approach that followed the stages of the eLearning Lifecycle (Phillips, McNaught, & Kennedy, 2011). This model was developed to specifically provide a framework for the development, implementation and review of electronic learning environments and, as such, was appropriate for the research focus.

The intended outcome of the eLearning Lifecycle is the identification of design principles to guide the design and development of future eLearning environments (Phillips et al., 2011). Based on the research study described here, several design principles were identified. Each is presented, in detail, together with key findings later in the paper.

The emergent functions of the ePortfolio in education

ePortfolio implementation has increased over recent years, driven largely by government policies that dictate their mandatory use (particularly in the United Kingdom) (Joyes, Gray, & Hartnell Young, 2010; McAllister, Hallam, & Harper, 2008). An ePortfolio is:

A collection of authentic and diverse evidence, drawn from a larger archive representing what a person or organization has learned over time on which the person or organization has reflected, and designed for presentation to one or more audiences for a particular rhetorical purpose (Barrett, 2005, p. 5).

This definition introduces the notion that a person may use an ePortfolio as a collection of documents and products for the purpose of evidence, possibly including personal records and reflections. However, the ePortfolio has clearly moved beyond this simple original function to facilitate the creation of a more purposeful and comprehensive presentation portfolio (Pelliccione, Dixon, & Giddings, 2005) that demonstrates a range of skills and abilities and that may be adapted to meet different purposes (Clark & Eynon, 2009).

A more comprehensive definition of the variety of ePortfolio affordances was elucidated by Duncan-Pitt and Sutherland (2006) who described it as:

A system that belongs to the learner, not the institution; populated by the learner not their examiner; primarily concerned with supporting learning not assessment; for life-long and life-wide learning not a single episode or a single course; that allows learners to present multiple stories of learning rather than just a simple aggregation of competencies; and, importantly, where access to them is controlled by the learner who is able to invite feedback to support personal growth and understanding. (p. 70)

With technological developments and the increased use of ePortfolios, a comprehensive range of functions has been identified. The key learning requirements that can be met through ePortfolios include: assessment, presentation, learning, personal development, collaboration, and ongoing working documents (Stefani et al., 2007, pp. 13-14).
There are three key aspects of ePortfolios that were particularly relevant to the current study—specifically assessment, learning and the scaffolding of complex tasks—which are discussed in detail in the following section.

The ePortfolio as an assessment tool

ePortfolios are being used increasingly as an assessment tool (MacEntee & Garii, 2010), particularly in terms of demonstrating skills, abilities and achievements against set criteria or required competencies (von Konsky & Oliver, 2012). A key advantage of the ePortfolio is that it can incorporate a variety of digital media that provides opportunities for a wide range of assessment submission formats, including the evidence of achievement against industry competency standards (Moran, Vozzo, Reid, Pietsch, & Hatton, 2013). An ePortfolio may also create a more authentic opportunity for students to demonstrate their learning (Raison & Pelliccione, 2006), and it allows for reflection in relation to learning goals and the planning of professional development (von Konsky & Oliver, 2012). While assessment is an important part of the learning process, using an ePortfolio for this purpose focuses specifically on the end product and may not fully explore the potential of the platform as a learning tool.

A useful extension to the understanding of ePortfolio environments was provided by Joyes and colleagues (2010, p. 2) who added that, “behind any product, or presentation, lie rich and complex processes of planning, synthesising, sharing, discussing, reflecting, giving, receiving and responding to feedback.” The encompassing term used by these authors to describe these processes was ePortfolio-based learning, which became the preferred term of the Joint Information Systems Committee (JISC) (Joyes et al., 2010). It was this extended focus on the ePortfolio as a learning tool that provided the next area of focus.

The ePortfolio as a learning tool

In 2005, Helen Barrett advocated that ePortfolios be utilised primarily as learning tools. She proposed that the technology was available to “engage students in active participation” using reflection by “assessing and managing their own learning” (Barrett, 2005, p. 23). As technology changes so rapidly, this statement is perhaps even more apt now.

A number of authors have written about the potential of ePortfolio platforms for teaching and learning (Housego & Parker, 2009; Lorenzo & Ittelson, 2005; Stefani et al., 2007). Throughout the United Kingdom, ePortfolios are widely in use, but the implementation process is still under development due to wide variation of opinion across the range of involved stakeholders (Joyes et al., 2010). Similar patterns appear to be occurring in the United States.

The collaborative work of the Australian ePortfolio Project (Hallam et al., 2010) and others have developed some important guidelines around the implementation of ePortfolios for teaching and learning. A review of these implementation initiatives reveals key recommendations for a successful environmental implementation. These include:

- Focus on the reason for the use of an ePortfolio (McCowan et al., 2005)
- Embed the implementation into coursework with a clear framework (Hallam et al., 2010)
- Scaffold the approach with both pedagogical and technical support (Pelliccione et al., 2005) and
- Allocate time to effect long-lasting change (Hiller et al., 2007).

Much of the existing research into the use of ePortfolios as learning tools in education has been specific to the platform being used (e.g., Mahara, PebblePad, WordPress). What was needed was a more universal approach with a focus on pedagogical considerations that could be applied across numerous electronic platforms (Shepherd & Hannafin, 2011). The ePortfolio can be used to complete a multitude of tasks, including those that require higher order thinking skills as discussed in the next section.

Using ePortfolios to scaffold complex tasks

When deciding to use an ePortfolio, it is important to first examine the task(s) to be completed within the environment and ensure that they are matched to the strengths and capabilities of the platform. For the ePortfolio in this study, the primary task for the students involved was the completion of an action research project. Action research involves the selection of an area that is of concern for the student in practice, and the subsequent design
and implementation of a cyclic approach of action and review to improve their abilities in this identified problem area (Grundy, 1995; McNiff, 1995).

The execution of such a research project is a complex and authentic task that requires students to self-regulate their learning performance by examining their own practice, comparing it to the literature and making positive change to their performance (McLoughlin & Lee, 2010). Through the completion of complex projects, students can assemble and display contributory products as they work towards a polished product, rather than assemble a range of possibly disconnected artefacts that demonstrate evidence of achievement (Herrington et al., 2010). An authentic task—in this case, the design, implementation and presentation of an action research project—provides opportunities for students to make decisions and reflect, as well as articulate and collaborate as they prepare their final publishable report. This is imminently achievable through the completion of a genuinely engaging and complex task supported within an ePortfolio environment.

The context of the study

This research study involved 4th year pre-service teachers in an Australian university. A compulsory unit in this final year required students to complete an action research project in an area of their choice. The implementation of this project required the students to access a classroom environment to trial and review the strategies under investigation as part of their action and review cycles. The project was scaffolded in the PebblePad ePortfolio through the provision of prompts to outline the assessment items and additional tasks to enhance reflective thinking.

Methodology

The implementation of this environment as part of the research study followed the cycles of the eLearning Lifecycle (Phillips et al., 2011). The model is based on the tenets of both Action Research, in terms of the cyclic implementation towards improvement in practice, and Design-Based Research that Phillips and colleagues (2011) believed provided a more structured research approach and focused on the creation of design principles.

The outline of the implemented Lifecycle, as shown in Figure 1, provides a framework of the stages implemented in the research and the phases of action and review that occurred at each point of the study.

![Figure 1. The eLearning lifecycle (Phillips et al., 2011)](image-url)
The ePortfolio was introduced to the students to complete the stages of their action research project and report on their results. The students were expected to continually use the ePortfolio to demonstrate their learning during the semester and document the project’s development. Within the PebblePad platform, the researcher provided prompts to the students via a Blog facility to scaffold the completion of the assessment tasks. The prompts in this Blog were designed to develop an ePortfolio-based learning environment to scaffold not only the assessment tasks, but also the development of reflective skills and dispositions through the stages of the action research project. The prompts were planned to meet the components of the Enculturation Teaching Model (Tishman et al., 1993) in the three areas of exemplars, activities and interaction. Although interaction was part of the wider study, the discussion in this paper is focused on exemplar and interaction prompts only.

The exemplar prompts were designed to demonstrate to students how to use the ePortfolio together with the frameworks required for the submission of assignments. These were detailed step-by-step instructions that guided the students through the stages of completing and submitting the assignment tasks within the PebblePad environment.

The activity prompts of the enculturation teaching model were designed to provide opportunities to practice higher-order thinking (Tishman et al., 1993). After examination of research into reflection, these activity tasks were developed based on strategies that had been found to be successful in enhancing reflective thinking and writing.

**Participants**

The students who accessed the learning environment were 4th year Bachelor of Education students completing a specialisation in either Early Childhood or Special Needs education. There were originally 84 students enrolled with access to the PebblePad Gateway that housed the scaffolding prompts. In total, 79 students completed the unit.

**Implementation**

The initial environment was implemented with a small group of students (see Round 1 in Figure 1) who were given prompts based on self-identified areas of need. In Round 1 most requests related simply to the assessment requirements of the unit and how to use the platform, rather than more complex, higher order learning issues. From this round of implementation it was decided that the activities needed to be pre-planned to meet the needs of students, but also to guide and prompt further exploration of the ePortfolio platform and to extend student thinking in the areas of their action learning project. Throughout the implementation of Round 2, the students were provided with the prompts outlined in Table 1.

The table outlines the number and name given to the prompt in the Blog, the source that was the basis of the idea for the prompt - either from literature or experience of Round 1 - and the purpose in the provision of each prompt. The shaded sections represent the prompts directly relating to the assessment tasks the students were required to complete.

Based on initial student feedback throughout this level of implementation, it was identified that while students utilised the prompts they did so to varying degrees. It emerged that the students accessed and implemented tasks mainly focused on the assessment tasks (the shaded prompts), but not those designed as higher-order activities or to enhance interactions and reflection.

<table>
<thead>
<tr>
<th>Prompt no.</th>
<th>Activity prompt</th>
<th>Source</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Reflection on teachers</td>
<td>Phillips &amp; Carr (2006)</td>
<td>The students were asked to describe and share what they think are the attributes of a good teacher. This was part of the action research project in terms of identifying something from their practice that needed improvement.</td>
</tr>
<tr>
<td>2</td>
<td>Something to talk about</td>
<td>Round 1 experience</td>
<td>This prompt was designed to encourage the students to use the blog for discussion.</td>
</tr>
<tr>
<td>3</td>
<td>Reflective Journal as a Blog</td>
<td>Spalding, Wilson, &amp; Mewborn (2002)</td>
<td>Reflective writing can promote reflective thinking because it is a permanent record of thinking, is an outlet for feelings, and can open up dialogue.</td>
</tr>
</tbody>
</table>

26
Table 2. Prompts for Round 2

<table>
<thead>
<tr>
<th>Prompt no.</th>
<th>Activity prompt</th>
<th>Source</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>Reflective writing review</td>
<td>QUT (DRAW Project), (n.d.)</td>
<td>This activity was a link to a platform for evaluating the level of reflection in student writing. Students could reflect on their writing against the determined criteria to make improvements.</td>
</tr>
<tr>
<td>10</td>
<td>Outline of 4R Framework Video Review</td>
<td>Ryan (2011)</td>
<td>This provided the graphic of the 4R framework mentioned in Prompt 9 for the students to use.</td>
</tr>
<tr>
<td>11</td>
<td>Verbal 3 step framework Reflective journal review</td>
<td>Jensen, Shepston, Connor, &amp; Killmer (1994)</td>
<td>The students were asked to video or audio record a teaching experience to review their practice with the review statements.</td>
</tr>
<tr>
<td>12</td>
<td>Verbal 3 step framework Reflective journal review</td>
<td>Donaghy &amp; Morss (2007)</td>
<td>The students completed a mini action research cycle on one event in their experience.</td>
</tr>
<tr>
<td>13</td>
<td>Conclusion questions</td>
<td>O'Connor &amp; Diggins (2002)</td>
<td>As the students began to put their projects together, this prompt aimed to get them to go back over their entries to add further detail or extra links to theory.</td>
</tr>
<tr>
<td>14</td>
<td>Final report</td>
<td>Round 1 experience</td>
<td>This assisted the students to bring their projects together and provided an overall review for their concluding chapter.</td>
</tr>
<tr>
<td>15</td>
<td>Final report</td>
<td>Round 1 experience</td>
<td>As with the Plan/Rationale and Progress Report, this prompt gave the guidelines for the assignment submission.</td>
</tr>
<tr>
<td>16</td>
<td>Attachments</td>
<td>Student request</td>
<td>This provided students with instructions to upload evidence and forms required with their submissions.</td>
</tr>
</tbody>
</table>

To address these issues, changes were made to the prompts provided. These included: (1) the addition of a source citation for the activity to demonstrate to the students that there was a theoretical basis to the prompts, and (2) the inclusion of questions to prompt discussion among the students. These changes were implemented (Round 3) and included the prompts listed in Table 2. The prompts most relevant for assessment tasks are again shaded for comparison.

At the conclusion of the implementation, students were invited to take part in the formal review process. This review employed a number of data collection methods including an online survey, focus group and individual interviews, together with the low-level utilisation of learning analytics and document analysis.

In all, 25 students responded to the online survey, 7 were involved in focus group interviews, 8 in individual interviews, 10 provided comments on the blog and another 10 provided feedback via email. As such, feedback in some form was received from over 50% of the cohort. The data from all these sources were then reviewed and analysed to identify possible improvements in the learning environment, and to create design principles for future iterations of ePortfolio-based learning environments.
Findings

To examine the students’ reported engagement with specific prompts, descriptive statistics were collated from the online survey. A question in the survey included a Likert ranking scale for the prompts that were used in the implementations and the students were asked to rank their level of engagement with each prompt as shown below:

- Did not look at
- Read only
- Read and used in project
- Read and completed activity
- Shared writing from activity with others.

The differentiation between “read and used in project” and “read and completed activity” was important to identify whether or not students were completing work unrelated to assessment items. The final category of “shared writing from activity with others” was to examine the use of the collaborative nature of the platform.

Table 3 presents the data collected from this question in the online survey. The numbers in the table represent the percentage of students who reported using the prompts at each level on the Likert scale question in the online survey.

<table>
<thead>
<tr>
<th>Prompt number</th>
<th>Prompt provided to students</th>
<th>Didn’t look</th>
<th>Read only</th>
<th>Read and used in project</th>
<th>Read and completed</th>
<th>Shared</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Reflection on teachers</td>
<td>46.7</td>
<td>46.7</td>
<td>6.7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>Something to talk about</td>
<td>33.3</td>
<td>60</td>
<td>6.7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>Reflective Journal as a Blog</td>
<td>26.7</td>
<td>53.3</td>
<td>6.7</td>
<td>13.3</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>Plan/Rationale Outline</td>
<td>6.7</td>
<td>13.3</td>
<td>66.7</td>
<td>13.3</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>Time to refine</td>
<td>40</td>
<td>40</td>
<td>13.3</td>
<td>6.7</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>Adding ethics checklist</td>
<td>6.7</td>
<td>6.7</td>
<td>60</td>
<td>26.7</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>Progress report</td>
<td>0</td>
<td>6.7</td>
<td>66.7</td>
<td>26.7</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>Uploading evidence</td>
<td>0</td>
<td>7.1</td>
<td>57.1</td>
<td>35.7</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>Reflective writing review</td>
<td>0</td>
<td>42.9</td>
<td>42.9</td>
<td>14.3</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>Outline of 4R Framework</td>
<td>6.7</td>
<td>33.3</td>
<td>60</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>Video Review</td>
<td>46.7</td>
<td>53.3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>Verbal 3 step framework</td>
<td>42.9</td>
<td>57.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>Reflective journal review</td>
<td>20</td>
<td>60</td>
<td>13.3</td>
<td>6.7</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>Conclusion questions</td>
<td>26.7</td>
<td>26.7</td>
<td>40</td>
<td>6.7</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>Final report</td>
<td>0</td>
<td>0</td>
<td>73.3</td>
<td>26.7</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>Attachments</td>
<td>0</td>
<td>21.4</td>
<td>42.9</td>
<td>35.7</td>
<td>0</td>
</tr>
</tbody>
</table>

The data shown in Table 3 highlights the increased usage of prompts that were provided as examples or directions for the completion of assignment tasks (the shaded sections). Based on the sum of the columns of “read and used in project” and “read and completed activity”:

- 80% (66.7+13.3) of students who completed the survey reported using or completing the Plan/Rationale activity from the outline.
- 93.4% (66.7+26.7) the Progress Report prompt.
- The full 100% (73.3+26.7) of the respondents (32% of the cohort) reported using or completing the Final Report submission from the exemplar prompts provided in the Gateway Blog.

This indicates that the students were fully engaged with the exemplar prompts that were related to the practical aspects of assignment submissions.

In contrast, the activity prompts that were provided to enhance the process of reflection were less frequently utilised. More specifically, the Reflective journal as a blog, the Time to refine, and the Reflective journal review were used by 20% of the respondents, while the Video review and Verbal 3 step framework were not used at all. The average use of the activity prompts was bolstered by the reported usage of the Reflective writing review (57.2%), the Outline of the 4R’s (60%), and the Conclusion questions (46.7%), which related specifically to reflective writing. This research finding is important because these prompts were directly related to the process of reflective writing that the students needed to complete for their assignment submissions. This finding places
these prompts at the intersection of process and product and reinforced the students’ preference for tasks related to the submission of assignments.

Responses from focus group and individual interviews were examined to further understand why students had utilised the prompts in this way. During the focus group interviews, several students identified that they completed only those elements associated with assignment tasks for a number of reasons: firstly, time constraints—the students had time to complete only tasks they felt were of greatest value for assessment; secondly, confusion about the purpose of the additional tasks; and thirdly, resistance to the use of a new platform that was imposed upon them in their final year.

**Time constraints**

Lack of time was mentioned repeatedly through the interviews as a contributing factor in the reduced engagement with the activity prompts. Many of students commented that they were “time poor” throughout the semester. Some said that even if the activity from the prompt had sounded interesting, they most likely would not have completed it unless there was an extrinsic value associated with it, in terms of their grade. One student described a somewhat cursory approach to completing the activities:

> I read some of them but I just – time was a factor. I did the first one but then there was too much else going on to do it. (Alex, Individual Interview)

While others said they completed some but not all due to time:

> Some of them I did, some of it I didn’t to be honest. It came down to my time and what I needed. (Monique, Focus Group Interview 3)

The non-assessment of the activity prompts appeared to be a real factor in their decreased use. Generally, the students did not make the link between the activity task in the prompt and their future teaching practice, but recognised that assessment results did impact on their course completions (Van Dinther, Dochy, & Segers, 2011). Assessment was clearly their primary focus.

Because the additional activities were not assessed within the unit, the students were not motivated to complete the task. Motivation is defined as “a complex psychological construct that attempts to explain behavior and the effort applied in different activities” (Watters & Ginns, 2000, p. 302). It cannot be imposed but must be actively chosen (Cohen, 1983), meaning that the students had to decide for themselves to make the activities a priority.

The difficulty engaging students in activities that are not assessed has been identified as a major issue in many online learning platforms (Macdonald, 2004). Macdonald stated, “the design of assessment is critical in determining the direction of student effort” (p. 218). While assessing every task in the ePortfolio is not sustainable (or desirable), it is important to examine how to encourage student engagement in tasks that are not assessed, leading to the identification of Design Principle 1: Utilise the ePortfolio in the weekly tasks (either online or on campus) to get the students using the tool as part of their regular work routines.

If regular use of the ePortfolio is embedded within weekly tasks, and each builds upon the resources and assets stored within this space, it may increase utilisation of the platform and encourage students to engage with the data collected within their pages. The students can continue to build these stores and access the required pieces for assignment submission, thus providing the link to assessment that has been found to be required (van Dinther et al., 2011).

**Purpose of the prompts**

Analysis of the interviews revealed that there was confusion regarding the purpose of the activity prompts. The discussion in the interviews highlighted that many students thought the prompts were provided for students who were struggling with the unit and needed additional help, rather than of being of assistance to the whole student cohort. For example:

> I think for someone who was really struggling in those aspects I would have used them but I didn’t feel like I had to write it down. (Kate, Focus Group Interview 2)
Students need to identify the long term benefits of using an ePortfolio platform (Edwards, 2013) and then be able to make the link between the completion of the task required and long term development. This led to the identification of Design Principle 2: Ensure the purpose of the task is clear and demonstrate to the students how completing it links with their learning.

In providing activities for the students, it is important that the process does not become one of direct instruction. Instead, the purpose must be made explicitly clear. The students need to identify “what’s in it for them” when completing the task as part of their development as self-regulated learners. By clearly outlining the purpose and demonstrating how useful the task may be in the future, they are then able to make informed choices about the value of task.

**Resistance to the technology**

The third area identified by students as a barrier to engagement was that the ePortfolio platform was a new and unknown program. PebblePad was introduced to the students in their final year of study and they were required to use it for only one subject. This was a key limitation of the research study and highlighted the need to introduce electronic learning platforms, such as ePortfolios, based on strong pedagogy from the commencement of their studies (Hallam et al., 2010). Although attempts were made to reduce the students’ discomfort with the platform, some resistance remained.

Several students highlighted negative perceptions of PebblePad and their preference for hard copies of documents as reasons for not engaging with the platform:

> Because I’d not used PebblePad before and I didn’t know what it was. I’d just heard horror stories from other students so I was a little scared. (Chloe, Individual Interview)

> I will do the hard copy cause I’m still learning a lot about technology as well so I think I still like doing hard copies. (Abigail, Individual Interview)

The format chosen as the assignment submission template was selected as it was the most flexible format in PebblePad and was the most similar to Microsoft Word. However, there were still small differences that may have presented challenges to students, as supported by the research of Janosik and Frank (2013). They found that students had difficulty adapting to format changes from those they were more comfortable with. This appeared to have been the case with the PebblePad environment, as many students reverted to the use of word processing or other familiar platforms to draft their work, which was then transferred to PebblePad.

The use of other platforms by students to complete the majority of the project work, and then transferring or “cutting and pasting” the contents over for submission, had a negative impact on their engagement with the ePortfolio platform. It reduced the amount of time students’ spent in PebblePad and, as such, decreased the depth of engagement with the platform.

It was not only the formatting of the platform that concerned students. Several voiced their frustration at having to learn another new and complex program towards the end of their degrees:

> It is hard enough other than adding in new programs every change of subject. To me this is wasting my time, it is hard enough to survive and take out time to do study, let alone having to learn new programs all the time. (Respondent 2-Online Survey)

Rather than seeing the new program as a useful tool, one student identified it as a “hurdle”:

> I found the use of PebblePad quite frustrating. I would have preferred to use PebblePad prior to my last year of uni rather than have another hurdle to try and overcome. (M-Email feedback)

The identification of factors that caused student frustration and resistance to the program emerged as an important consideration for anyone implementing an ePortfolio-based learning platform. This led to the development of Design Principle 3: Implement the ePortfolio with a strong pedagogical focus from the commencement of student degrees.

Linked with the first two design principles, this last one further highlights that the purpose of using the ePortfolio environment must be clearly embedded in a strong pedagogy. By introducing the platform to students
commencing their studies, individual ePortfolios can grow and develop alongside the students who create them throughout their degrees.

The findings of this research study provide some clear design principles for the ongoing implementation of ePortfolio-based learning environments. The goal in implementing these principles is to encourage students not only to create finished products, but also to become engaged in the process of developing reflective abilities.

Conclusion

The aim of this research study was to investigate the use of an ePortfolio platform as a learning environment to scaffold the enhancement of reflection in pre-service teachers. While there were some successful outcomes of this implementation particularly in relation to development of students’ self-reporting of reflective skills, a number of design principles emerged to guide future iterations of these environments. The cyclic implementation of the eLearning Lifecycle showed:

- ePortfolios can be used to provide regular reflective tasks as an integral part of students’ learning routines,
- The purpose for each task needs to be made clear to students, so they can appreciate the value of completing the work,
- For maximum benefit, there should be a strong pedagogy behind the implementation of the ePortfolio from the commencement of the students’ degree programs, rather than an initiative introduced late in a program of study,
- Opportunities should be provided for the students to decide what could be included in their ePortfolios to achieve their learning goals.

These principles highlight the importance of the learning tasks implemented within an ePortfolio-based environment and the need for a strong pedagogical approach to the use of such platforms. The activities required of the students must engage them at a complex level and require multiple layers of scaffolding to be implemented—and acted upon—by the students within the environment.

These design principles can be applied to any ePortfolio platform for a multitude of task types and content areas. It is anticipated that the application of these design principles will meet the needs of the students and, in this way, increase the efficacy of ePortfolio-based learning environment implementation to better scaffold complex tasks for student learning and development.

This study also suggests at least two directions for future research. Firstly, this model could be re-implemented in a similar situation to examine design principles used in a different context. This would add to the evidence of the effectiveness of the recommendations. Secondly, further studies could build on these principles in terms of the collaborative affordances of ePortfolios, and investigate strategies to increase the use of these functions amongst students creating assets in collaboration.

References


An Analysis of Density and Degree-Centrality According to the Social Networking Structure Formed in an Online Learning Environment

Esin Ergün1* and Yasemin Koçak Usluel2

1Computer Technology Department, Karabuk University, Turkey // 2Computer and Instructional Technology Department, Hacettepe University, Turkey // esinergun@karabuk.edu.tr // kocak@hacettepe.edu.tr

*Corresponding author

(Submitted April 5, 2015; Revised April 1, 2016; Accepted April 18, 2016)

ABSTRACT

In this study, we assessed the communication structure in an educational online learning environment using social network analysis (SNA). The communication structure was examined with respect to time, and instructors’ participation. The course was implemented using ELGG, a network learning environment, blended with face-to-face sessions over a 14-week period. Data were collected from 114 undergraduate students who were enrolled in Instructional Technology and Material Design course. The program functions on the basis of a matrix; in this case a square matrix with rows and columns being the students’ ID numbers. Density and centrality measures were visualized and interpreted. In terms of the density of the groups, it was found that the lowest density occurred during the first week. The highest density, on the other hand, occurred during the week when the instructor participated, in all the groups except for the third and sixth groups. The students placed in the center and those on the edges of the network differed on the basis of time as well as the instructor’s participation. Other online learning environments could be assessed in a similar fashion using SNA in order to understand levels of participation and changes in interaction over time.

Keywords

Social network, Density, Centrality, Social network analysis

Introduction

Online learning environments may be inadequate in terms of establishing collaboration, giving feedback and receiving social support - which are relatively easier to achieve in traditional educational environments; this may influence the interpersonal relations in the learning environment. It is an indispensable component for the new generation of students, labeled as digital natives, to share by using innovative technological instruments such as social networks, blogs - to obtain and to create knowledge - and maintain high levels of communication. Social networking relies on the relationships students form among each other. Social networks become conduits for information, knowledge sharing and much more – it also affects members’ behaviors (Grunspan, Wiggins & Goodreau, 2014).

Using social networks in educational and instructional contexts can be considered as a potentially powerful idea simply because students spend a lot of time on these online networking activities (Mazman & Usluel, 2010). For example, social networks facilitate the process of communication between students and teachers, ensure their participation, offer peer support, and support cooperative learning. Today, teachers are able to re-shape the learning-teaching process through social networks formed in online learning environments by means of Web 2.0 technologies. Both synchronous and asynchronous communication is possible in these online environments where learners and educators communicate using such tools as forums and blogs. It would be both difficult and time-consuming for educators to analyze those messages manually (Rabbany, Takaffoli, & Zafane, 2011). The statistical analyses of those messages have informed us only the frequency of messages between each other until recently. Yet, there was the likelihood of making mistakes in those analyses as well and these analyses occasionally failed to inform us about learner-teacher and learner-learner communication. The communication patterns of students who are in the center position as well as their social behaviors can be analyzed by means of SNA. The rate of knowledge sharing through online communication can be identified in this way, while the significance of the patterns that emerge can be analyzed. Students in key roles can affect the flow of communication, and influence the amount of information and communication, thereby affecting interpersonal cooperation in a teaching and learning environment. Therefore, giving important responsibilities to key actors in a group can make learning more effective, but only if the instructor understands the characteristics of the network formed by students.
SNA is employed in order to study social networks, i.e., to reveal the multi-relations between organizations, employees, clients and students. Network data are used to find out how knowledge is transferred between individuals or to test the efficiency of organizations in terms of knowledge flows. Cela, Sicilia and Sánchez (2015) state that, when applied to learning activities, SNA usually aims to identify factors that influence the success or efficiency of the educational process. SNA-related research studies are seen to have found their way into educational environments since they help to identify the patterns between individuals who are part of the same social network such as students in the same class. This is due largely to data that enable research to become knowledgeable about the importance of social communication in online learning environments. Amongst others, network characteristics affect learners’ achievement (Penuel et al., 2006; Cho et al., 2007; De Laat et al., 2007; Grunspan, Wiggins & Goodreau, 2014). The relations among members interacting with each other in a social circle can be mapped (Wasserman & Faust, 1994). Those patterns that emerge can help us to model the flow of information between participants. De Laat et al. (2007) demonstrates what can be done with SNA in learning environments as follows:

- Since group behaviors are not static, researchers can find answers through SNA to the questions about relationships such as students-students and students-teachers and how these affect learning in learning environments and how they develop their competencies.
- Patterns of interaction on how the flow occurs between members can be revealed. While some of the participants move towards the center of the network, some others get away from the center. Thus, it can clearly be seen who manages the discussion and who is dominant.
- Relations amongst participants in a network, both direct and indirect, can be visualized, measured and analyzed.
- When used in combination with other methods such as content analysis and interviews, SNA can offer detailed information on learning and teaching processes.
- It enables us to monitor the changing relations between group members, the contributions made, and the experiences.

Social networks consist of actors and their relations. These are indicated as nodes while ties bind those nodes that have a relationship. In the case of an online learning environment, the communication flow between students (nodes or actors) is indicated by ties (links or edge) between those nodes that communicate. The network is the combination of a series of ties between students or a set of relations between them (Haythornthwaite, 2005). The students’ positions in the network can be determined by means of the measurements performed with SNA and it is possible to demonstrate the way their positions affect their learning environment. The measurements that are most frequently used in research studies are degree, betweenness and closeness centrality, which are analyzed within the scope of density and centrality.

SNA is an important method of analysis for determining students’ patterns of communication and interaction, for raising efficiency and for ensuring continuity. In short, it is possible to state that SNA offers alternatives distinct from traditional research methods and more realistic in evaluating communication patterns among students, especially in online learning environments given the ease with which network data can be obtained. To the best of our knowledge, there is a limited number of studies that analyze the structure of the social network in educational settings. In previous research, SNA was used to examine how learner networks evolve over time (De Laat et al., 2007), to examine using communication channels to achieve their educational goals (Haythornthwaite, 1999), to create patterns of relationships connecting teachers in networks (Ryymin et al., 2008), to determine the position of individual students’ influence in learning outcomes (Cho et al., 2007). Consequently, it may be said that social networks are the important components in regulating the online learning environments. Based on this fact, this study aims at examining how students’ social network structures change over a set period of time in an online learning environment. For our purposes, answers are sought to the following questions:

- What are the density measures of student groups?
  o How does the group of students’ density change on the basis of time?
  o How does the group of students’ density change on the basis of the instructor’s participation?
- What is the students’ degree centrality?
  o How does students’ degree centrality change on the basis of time?
  o How does students’ degree centrality change on the basis of the instructor’s participation?
Methodology

Participants

The Instructional Technology and Material Design course was offered over a 14-week period using ELGG, which is an online learning environment that augments contact sessions. Data were collected from 114 undergraduate students who were enrolled in this course offered by the educational faculty of a foundation university located in Ankara. 89% (n = 102) of the participants were female whereas 11% (n = 12) were male (Table 1).

Table 1. The number of students according to gender and groups

<table>
<thead>
<tr>
<th>Group</th>
<th>Gender</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1 – Pre-school Teaching</td>
<td>Female</td>
<td>24</td>
<td>21.05</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>23</td>
<td>-</td>
</tr>
<tr>
<td>Group 2 - Elementary School Teaching</td>
<td>Female</td>
<td>20</td>
<td>19.30</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>22</td>
<td>-</td>
</tr>
<tr>
<td>Group 3 – Elementary Mathematics Teaching</td>
<td>Female</td>
<td>7</td>
<td>9.65</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>4</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>11</td>
<td>-</td>
</tr>
<tr>
<td>Group 4 – Mixed (Pre-school and Elementary</td>
<td>Female</td>
<td>22</td>
<td>19.30</td>
</tr>
<tr>
<td>Mathematics Teaching)</td>
<td>Male</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>22</td>
<td>-</td>
</tr>
<tr>
<td>Group 5 – Turkish Teaching</td>
<td>Female</td>
<td>14</td>
<td>14.92</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>17</td>
<td>-</td>
</tr>
<tr>
<td>Group 6 – English Teaching</td>
<td>Female</td>
<td>15</td>
<td>15.78</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>18</td>
<td>-</td>
</tr>
</tbody>
</table>

Data analysis

The methods for mathematical measurement and calculation used in identifying and analyzing the social network differ from the ones used in the analysis of any quantitative research in social sciences. Some measurements, such as degree centrality, between centrality are employed in the SNA in describing the students’ positions in a network, in identifying the inter-student relations, and in determining a network formed as a whole. The SNA is used so as to analyze the structures of any type, which are in relation to one another or to the institutions, individuals or groups within the social structure (Wasserman & Faust, 1994; Freeman, 2004).

UCINET 6.0, one of the most well-known and the most frequently used software application when performing SNA, was also used in this study. The program functions on the basis of matrix, and it has a text-based format. The NetDraw program was used, however, for the visualization of the network.

Determining the actors’ roles through SNA is important in understanding the effects of the roles on the network structures. In the present study, the structure of social network was analyzed through the measurements of density and degree centrality.

Density: Density is defined as the proportion of ties existing in the social network to all probable ties (Borgatti, 2003). It displays the frequency of information flow between individuals. The density or the scarcity of the ties emerges as a property of the network. A dense network is a network in which the number of ties is close to the maximum. A network with small number of ties is called scarce. The density of a network is calculated by dividing the number of ties in the network into the probable number of ties available in case the network is a full network. Consequently, density shows the percentage of the ties used which are potentially usable (Gürşakal, 2009). It receives values between 0 and 1 in the binary number system. The 0 value demonstrates that there are no ties between students while 1 or above value shows the number of ties through which a student communicates with other students. As a unit of density, 100% demonstrates that each individual has talked to all individuals at least once (Lowes, Lin & Wang, 2007).
Centrality: Centrality measures the importance of students within a network, and shows which students are in the center (Borgatti & Cross, 2003). It provides information on the position of students in the network. The individuals in the center within the network are called stars. They are the most popular people in the network. All network centralization measures may range from 0 to 100%. 100% as the unit of centrality means that all participants talk to one single individual (Lowes, Lin & Wang, 2007). Degree centrality, betweenness centrality and closeness centrality are the commonly used centrality measures.

Degree centrality shows the actors’ degree of direct ties with the others in the network. An actor may be tied to another actor in linear relations in the ties. Generally, the higher is an actor’s number of ties, the more important and powerful he is. The actor with the highest degree centrality can be said to be the most active actor in the network. The measurement of degree centrality can help to identify the active participation of the key characters in an online discussion (Kale, 2007).

Data collection

Data obtained from the selected online course provided the matrix necessary for analysis using UCINET 6.0. However, the online learning environment used in the research did not enable us to obtain the data matrix directly. Therefore, the data matrix was derived from the environment through SQL code that provided the researchers with the required matrix data. The required matrices with this SQL code were then created automatically in Excel prior to importing into UCINET. This way, the necessary matrices for analysis were obtained automatically.

The matrices to be formed with the data coming from the network-learning environment are needed for the SNA. Those matrices can be used as the source of data and analyses can be performed on the UCINET 6.0 program. Adjacency matrices are usually used for the data in SNA. This stems from the fact that adjacency matrices show who is close to whom or which students are adjacent to another students’ social field. The students in the network are placed in the columns and lines of the matrices, and the number of ties between students is placed in the lines and columns in the binary or the decimal number system. The number of ties is placed in the matrices in binary or decimal number system according to the properties of the variable analyzed. Whether there are relations between students is to be revealed, the matrices encoded as 1; and otherwise it is encoded as 0. The density of student groups was calculated through binary adjacency matrix and the centrality was calculated through decimal adjacency matrix based on students’ number of ties. An example of decimal and binary adjacency matrix of research data are given below.

<table>
<thead>
<tr>
<th>A97</th>
<th>A103</th>
<th>A96</th>
<th>A98</th>
<th>A94</th>
<th>A95</th>
<th>A93</th>
<th>A104</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>7</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A97</th>
<th>A103</th>
<th>A96</th>
<th>A98</th>
<th>A94</th>
<th>A95</th>
<th>A93</th>
<th>A104</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

The discussion data for the six sub-groups formed in a network-learning environment in a 6-week period constitute the research data. A total of 36 data matrices were formed by using the data. The students’ names were encoded by systematically. Students’ names were encoded with alphabetical letters (A, B, C, D, E, and F) and the
instructor was coded with (G). Then, the analyses were performed by using those codes in the research in order to ensure anonymity.

Findings

Based on the research problems, the findings are listed as in the following:

How is the density of student groups?

*How does the group of students’ density change on the basis of time?*

The density of the student groups was calculated according to weeks (Table 2). García Hernández and Reyes López (2009) reported that the middle level value of density measurement is between 40% and 70% in SNA. These values are interpreted taking into account the density of the group.

<table>
<thead>
<tr>
<th>Group</th>
<th>Week1</th>
<th>Week2</th>
<th>Week3</th>
<th>Week4</th>
<th>Week5</th>
<th>Week6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.34</td>
<td>0.53</td>
<td>0.42</td>
<td>0.69</td>
<td>0.60</td>
<td>0.71</td>
</tr>
<tr>
<td>2</td>
<td>0.20</td>
<td>0.36</td>
<td>0.63</td>
<td>0.62</td>
<td>0.49</td>
<td>0.71</td>
</tr>
<tr>
<td>3</td>
<td>0.62</td>
<td>0.62</td>
<td>0.62</td>
<td>0.60</td>
<td>0.76</td>
<td>0.69</td>
</tr>
<tr>
<td>4</td>
<td>0.17</td>
<td>0.29</td>
<td>0.48</td>
<td>0.52</td>
<td>0.51</td>
<td>0.45</td>
</tr>
<tr>
<td>5</td>
<td>0.16</td>
<td>0.50</td>
<td>0.45</td>
<td>0.76</td>
<td>0.56</td>
<td>0.56</td>
</tr>
<tr>
<td>6</td>
<td>0.25</td>
<td>0.30</td>
<td>0.43</td>
<td>0.56</td>
<td>0.63</td>
<td>0.61</td>
</tr>
</tbody>
</table>

Group 1: It is clear that the density for the first group was the lowest (0.34) in the first week whereas it was the highest (0.71) in the sixth week. Accordingly, it may be stated that a low level of communication occurred in the first week and a high level of communication occurred in the sixth week in the discussion environment in which there were eight students.

Group 2: The density for the second group was the lowest (0.20) in the first week whereas it was the highest (0.71) in the sixth week. It is remarkable that there is a continuous increase until week five, followed by a decrease in week five, and then it reaches the highest level in week six. The density of communication in this group has increased continuously until the fifth week, but fell in the fifth week. Face-to-face course couldn’t be conducted this week due to national holidays.

Group 3: The density for the third group was the lowest (0.60) in the fourth week whereas it was the highest (0.76) in the sixth week. It is evident that the density of communication is very close in weeks apart from week six. There were twenty-five students in first week while there were twenty students in fourth week. Accordingly, it is possible to say that in all weeks of intensive message exchange occurred in this group.

Group 4: The density for the fourth group was the lowest in the first week whereas it was the highest in the fourth week. It was observed that the density of communication increased until week four, but it decreased after week four. It was 0.17 for the first week while it was 0.52 for the fourth week. Accordingly, it is possible to say that in general the message of the change at least the whole week in this group that consist of two different groups of students.

Group 5: It is clear that the density for the fifth group was the lowest (0.16) in the first week whereas it was the highest (0.76) in the fourth week. There were nineteen students in first week while there were eleven students in fourth week. Instructors have attended the fourth week in this group. It is noteworthy that after this week, the density of communication was felt away.

Group 6: The density for the sixth group was the lowest (0.25) in the first week whereas it was the highest (0.63) in the fifth week. There were nineteen students in first week while there were fourteen students in fourth week. Instructors have attended the fourth week in this group as same as group 5. Contrary to the group 5, density of communication increased after this week.

On examining the density of all groups in general, it was found that the lowest density was in the first week in all the groups except for group three while the lowest density for group three was in the fourth week (Figure 1).
How does the group of students’ density change on the basis of the instructor’s participation?

While the instructor participated in the discussion held in week six in the first, second and third groups; the instructor did not participate in the discussion held in week four in the fourth, fifth and sixth groups. The density was at the highest level in all groups apart from group six in the weeks when the instructor participated. The highest level for group six was in week five - a week after the participation of the instructor. Besides, it was also remarkable that the density was higher in week six than that in week four.

It was also remarkable that the highest density in groups apart from group six was in the weeks when the instructor participated. In group six, however, the instructor participated in the fourth week, and the highest level of density was in the following week - that is, in the fifth week. The density of group was higher after week four - the week of instructor’s participation - than in the previous weeks.

What is the students’ degree centrality?

The number of students’ ties with the others was taken into consideration in calculating the degree centrality. It was stated that the student with the most ties had the highest degree centrality whereas the student with the fewest ties with others had the lowest degree centrality. Accordingly, the student with the most ties with others is in the centre while the one with the fewest ties is outside the network.

How does students’ degree centrality change on the basis of time?

The results of analysis performed through SNA in relation to the students’ degree centrality on the basis of time are shown in Table 3. It shows only the number of student in the center of the network. Some weeks while one student place in the center of the network, some weeks there were more than one student in the center of the network. Accordingly, it is clear that the network maps are complex and dense with some groups whereas they are scarce with some other groups. One of the basic reasons for this is the number of students participating in the environment in that week. As is seen from the examples, while there were 8 students in group six; in week six, there were 25 students in group four in week one. The network maps showing the degree centrality for all groups by weeks are visible in Figure 2, Figure 3, Figure 4, Figure 5, Figure 6, and Figure 7.

According to the degree centrality measurement results, the students with high degree centrality and the students with low degree centrality in the groups are shown in the network maps (sociograms). The students marked in red are the students with the highest degree centrality in the network while the ones marked in blue are the students with the lowest degree centrality.
Table 3. The students with high degree centrality on the basis of time

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Theme</th>
<th>Weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Group 1:</td>
<td>Number of students</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Percentage of centralization of the network</td>
<td>67%</td>
</tr>
<tr>
<td>Group 2:</td>
<td>Number of students</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Percentage of centralization of the network</td>
<td>67%</td>
</tr>
<tr>
<td>Group 3:</td>
<td>Number of students</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Percentage of centralization of the network</td>
<td>47%</td>
</tr>
<tr>
<td>Group 4:</td>
<td>Number of students</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Percentage of centralization of the network</td>
<td>59%</td>
</tr>
<tr>
<td>Group 5:</td>
<td>Number of students</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Percentage of centralization of the network</td>
<td>87%</td>
</tr>
<tr>
<td>Group 6:</td>
<td>Number of students</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Percentage of centralization of the network</td>
<td>65%</td>
</tr>
</tbody>
</table>

Group 1: According to the Figure 2, a student (A96) has the highest degree centrality for all weeks. In week five, two students in addition to that student were also placed in the center, and two more students were added to them in week six. Although the number of participation in the discussion environment was small in week six, all except two students actively sent messages to one another, which is remarkable. On seeing the network as a whole, the network centrality values were found as 67.95% for the first week, 45.45% for the second week, 58.97% for the third week, 38.89% for the fourth week, 36.11% for the fifth week, and 19.05% for the sixth week. It may be said that centrality was not available in the fourth, fifth and sixth weeks.

Figure 2. The network maps for the Group1

Group 2: While there were two students with the highest degree centrality in weeks one and three, there was only one student with the highest degree centrality in the other weeks. It was found that some students sent messages only to their groups in the first week. On seeing the network as a whole, the network centrality values were found as 67.14% for the first week, 54.76% for the second week, 35.42% for the third week, 35.71% for the fourth week, 33.33% for the fifth week, and 42.86% for the sixth week. It may be said that centrality was not available in the third, fourth and fifth weeks.
Group 3: On examining the degree centrality by weeks, it was found that different students achieved high degree centrality each week. While only one student attained the highest score in weeks one, two and four; more than one student did so in the other weeks. In week six, two students who had not attained high degree centrality before were added to the students who had attained the highest degree centrality scores in at least one of the first five weeks, and thus five students attained the highest scores in week six. On seeing the network as a whole, the network centrality values were 46.67% for the first week, 34.55% for the second week, 34.44% for the third week, 48.89% for the fourth week, 30.56% for the fifth week, and 25.56% for the sixth week. It may be said that centrality was not available in the second, third, fifth and sixth weeks.

Group 4: On examining the degree centrality across the weeks, it was found that the same student (D2) always attained the highest degree centrality for six weeks. It was remarkable in the second week that a group was formed and messages were sent within the group only. On seeing the network as a whole, the network centrality values were found as 59.33% for the first week, 59.06% for the second week, 39.71% for the third week, 47.37% for the fourth week, 30.03% for the fifth week, and 38.95% for the sixth week. It may be said that centrality was not available in the third, fifth and sixth weeks.
Group 5: On examining the degree centrality by weeks, it was found that there was one student (E66) attaining the highest degree centrality four times. While only one student attained the highest degree centrality in the first five weeks, three students did so in the sixth week. It was remarkable that several students set up ties with only one student in the network structure formed in the first week. On seeing the network as a whole, the network centrality values were found as 87.25% for the first week, 42.5% for the second week, 55.13% for the third week, 41.82% for the fourth week, 51.28% for the fifth week, and 18.18% for the sixth week. It may be said that centrality was not available in the sixth week only.

Group 6: On examining the degree centrality by weeks, it was found that there was one student (F74) attaining the highest degree centrality four times. While only one student attained the highest degree centrality in the first four weeks, two students did so in the fifth and sixth weeks. It was remarkable that the number of students participating in the discussion decreased in the sixth week. On seeing the network as a whole, the network centrality values were found as 65.03% for the first week, 81.32% for the second week, 49.45% for the third...
week, 42.31% for the fourth week, 34.62% for the fifth week, and 50% for the sixth week. It may be said that centrality was not available in the fifth week only.

In summary, the network centrality was the highest for all groups in the first week. It may be said that the distribution of students within the network was not very high in the weeks with low percentages of network centrality. It was remarkable that there were not groupings in two groups except for two weeks.

How does students’ degree centrality change on the basis of the instructor’s participation?

The instructor participated in the first, second and third groups in week six and in the fourth, fifth and sixth groups in week four. The students’ degree centrality scores were examined for those weeks. There were five students attaining the highest degree centrality in the first group. Only one student attaining the highest degree centrality was available in the second group. That student participated in the environment in the other weeks, but could not attain the highest score. In the third group, there were three such students. This was the group with the most students in the center of the network. While one of those students did not participate in the environment apart from that week, the other student remained outside the network in another week. There was one student attaining the highest degree centrality in the fourth group. It was found that that student attained the highest degree centrality in the fifth and sixth weeks also. There was one student attaining degree centrality in the fifth and sixth groups. These students attained the highest degree centrality some weeks or had achieved a high degree centrality other weeks. In conclusion, according to the results of degree centrality, the number of students placed in the center of the network increased in the weeks when the instructor participated while the number of students remaining outside the network decreased.

Discussion

This study maps the relations among students enrolled for the Instructional Technology and Material Design course by using SNA. In line with our purpose, the relational data were collected via the forum formed in the online learning environment. The data were then transformed into sociometrical matrices in a statistical program performing the SNA. In accordance with the research problems, the density and centrality measures were used, and the data were visualized and interpreted.

Density measurement provides a prediction of the diffusion rate of knowledge between actors. Martino and Spoto (2006) point out that density is an indication of homogeneity of the group and actors’ engagements to each other. Therefore, if individuals who have a high density leave the online learning environment, the lack of communication network and the flow of knowledge can also be said to slow down in the group. On examining the density of the groups in this study, it was found that the lowest density was in the first week. The highest
Density was, however, in the other groups apart from the third and sixth groups in the week when the instructor participated. In this sense, Abedin, Daneshgar and D’Ambra (2014) also point out that the availability of an administrator in the environment supports students in terms of participation. It could be said that these students want to show themselves for having a good grade. When the number of students’ messages were examined at the weeks of instructor’s participation, the number of sending messages was greater than the number of receiving messages all of the groups. Yet, it may be said that the participation of the instructor did not affect the participation of the students in the discussion in both groups (group three and group six). An, Shin and Lim (2009) point out that students tend to be more comfortable in expressing their thoughts when instructors participate less to the online environment. On examining the time-dependent density, this case showed that either the density was high or it was low in all weeks in these two groups. Abedin, Daneshgar and D’Ambra (2014) state that facilitators can benefit from the students’ non-task behaviors encouraging other students to behave in a similar fashion. For example, as a result of an association between reflections, thanks, salutations, and signatures, if the facilitator encourages students to share personal and work experiences with others, he or she may start the discussions with posting his or her own experiences. This, in turn, will encourage others to make similar posts.

The students’ degree centrality scores were examined according to groups, and it was found that there were students in the center of and outside the network. The students placed in the center and remaining outside the network differed on the basis of time and of the instructor’s participation. In the week when the instructor participated in the environment, the number of students remaining outside the network decreased. Phang, Kankanhalli and Sabherwall (2009) stated that individuals considered the existence of a moderator in information seeking behaviors more important than in the behaviors of contributing to knowledge. Gómez, Roses and Farias (2012), on the other hand, state that teachers were important in the academic use of social networks and in ensuring students’ participation on checking the general network centrality values for all groups by weeks, it was found that it was in the 30%-60% range. This shows that students were not very active in the network. It may be said that the number of ties each student has in the network is close to each other. In brief, the students actively participating in the forum in the online learning environment, the ones leading the group in ideas and the ones with little importance in the group were determined with this measurement.

Degree centrality can be a measurement to determine the opinions of leaders. Rogers (2003) stated they are at the center of interpersonal communication networks. A communication network consists of interconnected individuals who are linked by patterned flows of information. It is important to adopt new practices and ideas within the group in order that opinion leaders take place in the online learning environment for ensuring the continuity of efficient resource sharing or discussion and raising the group dynamic. Opinion leaders communicate with others, develop social networks, manage and enable participation by communicating and helping. It is important for online learning communities to benefit from their strong and influential relationships with others in identifying, recognizing and motivating this people. It seems that the students at the center in groups varied according to the weeks. The number of messages sent by these students was determined to be greater than the number of messages received from others. Also, they generally started discussion with others. It can be said that they motivated others for participate to environment. In addition, it is observed in the network map that there wasn’t any clustering all of the groups. This indicates that students are interconnected, be it directly or indirectly. This situation brings about the question of whether the effect of the course carried out both online and face-to-face environment.

According to the results obtained within the present study, the following recommendations are made:

- **SNA** is a powerful tool in measuring the interaction occurring in the learning process in the online learning environments (Rienties, Tempelaar, Bossche, Gijseelaers & Segers, 2009). Even if only density and degree centrality were used in this study, different social network measurement would be used for to different purposes in online learning environments.

- **Students’ network centrality has correlated with learning achievement** (Lin, Huang & Chuang, 2015). Students with high centrality outperform students with low centrality on learning achievement. Stepyan, Borau and Ullrich (2010) state that students showing higher reciprocal interaction also showed higher achievement scores. Another research suggests that the position in the network is positively related to learning performance in computer supported collaborative learning (Cho et al., 2007). Marcos et al. (2016) points that a positive correlation between students’ performance and six of the metrics employed (degree, eigenvector centrality, betweenness centrality, hub, authority and PageRank). So, there will be a new study done about learning achievement with SNA.
• Besides increasing the density in online learning environments, the quality of the shared knowledge is also important. Therefore, the elements raising the quality rather than the quantity of the messages should be included. As the quality of the messages increases, it is recommended that experimental studies concerning how students’ engagements changes should also be performed.

• This study analyzed all of the response messages. The system logs were used, so each response message, which includes all written message by students and teachers, can be task related or non-task related. For example, it contains encouraging words or chitchat words. A new study will be done only by excluding these types of non-related task messages used in combination with other methods such as content analysis or interviews in the future.

• Gladwell (2008) stated that “we are too much in awe of those who succeed and far too dismissive of those who fail.” A need for studies investigating the content shared between both students in center and outliers, and the way in which they are shared ought to be beneficial to instructors and course designers.

References


ABSTRACT
The present study investigated the effects of providing subtitles and taking enotes on cognitive load and performance. A total of 73 English-as-a-Foreign-Language (EFL) undergraduates learned brain anatomy and cognitive functions through multimedia programs. We used a 2 (subtitle/no) x 2 (taking enotes/no) factorial design to test the following: (1) if cognitive demand would vary among different groups, and (2) if providing subtitles or taking enotes would facilitate learning. Four versions of multimedia materials had exactly the same content and differed only in the availability of subtitles or enote functions. The enote tool was designed for effective note-taking with regard to video/animation, making it possible to directly link notes to a specific scene. The results indicated that animation with subtitles groups help reduce cognitive load and increase performance. No significant differences were found in either cognitive load or post-performance for those who learned with or without taking enotes. Although previous studies suggest subtitles may lead to a split-attention effect for native learners, we found a dual-coding effect (or reverse split-attention effect) for EFL learners.

Introduction
Learning science concepts with the aid of multimedia can facilitate learning, especially for complex learning content (Mayer, 2009; Mayer, 2011; Mayer & Moreno, 1998; Mayer & Moreno, 2002). Therefore, multimedia with animations has been widely used to support learning in domains such as earth science (Chang & Yang, 2010), physics (Mayer & DaPra, 2012; Mayer & Moreno, 2002), and chemistry (Chang & Yang, 2010). Most of the success in multimedia learning can be well explained using the working memory model (Baddeley, 1992), which asserts that there are two separate but interconnected audio and visual channels for processing speech-based information and visual images, and Paivio’s dual-coding theory (Clark & Paivio, 1991), which distinguishes between two semantic memory systems, one for pictures and another for words. Learning multimedia materials with both narration (spoken text) and subtitles (written text) can cause a split-attention effect (Mayer, Hegarty, Mayer, & Campbell, 2005). This phenomenon is consistent with cognitive load theory (Chandler & Sweller, 1991) in that that mental integration of written text and graphics will overload working memory and hamper learning. These studies mainly focus on multimedia learning using learners’ native language. Few studies have researched issues in an EFL learning context. Some research has provided positive evidence regarding the effect of subtitles for EFL learners. For example, Mitterer and McQueen (2009) suggest that foreign subtitles were beneficial for foreign learners in creating lexical interference.

In the age of Web 2.0, an increasing number of science materials are shared (e.g., TED Talks), and online courses (e.g., MOOCs) are easy to access. Therefore, teachers are very excited about introducing science materials originally produced with English to their EFL students. However, scientific knowledge involving abstract concepts is not always comprehensible, especially in a foreign language context. Fortunately, thousands of volunteers have contributed to translating and editing on-screen subtitles (e.g., TED Talks) based on the belief that translated subtitles would promote the programs and the understanding of complex scientific concepts for non-native users. Does viewing subtitles really ease cognitive load and increase understanding when the language is not the users’ native language? Or are subtitles an extra source of cognitive load that can possibly split non-native users’ attention while attending to visual information such as animation? On the other hand, over the past decades, research on multimedia learning has shifted from media for information delivery to a student-centered paradigm in which learners control their pace of learning. People can stop, resume, or review multimedia materials (e.g., videos) as many times as they need. Nevertheless, limited working memory still poses an obstacle to their learning processes. Note-taking in the multimedia learning environment may serve as an extension of working memory to assist students with mentally processing learning materials (Newton, 2000).
Cognitive load theory in multimedia learning

Cognitive load theory has been adopted to interpret the load on humans’ cognitive system in performing a task (Chandler & Sweller, 1991; Mayer, 2005). Instructors apply cognitive load theory as a framework in lesson planning (Paas, Renkl, & Sweller, 2003a). A basic assumption is that the human cognitive system, including attention and working memory, is limited. Therefore, complex tasks are more likely than simpler tasks to overload the system. Overall cognitive load consists of three components, namely, intrinsic cognitive load, extraneous cognitive load, and germane cognitive load (Paas, Tuovinen, Tabbers, & van Gerven, 2003b). Intrinsic cognitive load is mainly caused by the number of elements that must be processed simultaneously, in addition to the interrelationship between elements (Sweller, Ayres, & Kalyuga, 2011). Extraneous cognitive load results from inefficient instructional designs that are demanding for learners (Park, Moreno, Seufert, & Brünken, 2011). Germane cognitive load is the load when a learning task is acquired with or integrated into previous schema (Paas & Van Gog, 2006). The three types of cognitive load are assumed to be additive, and learning can only be achieved when the sum does not exceed the overall cognitive capacity (Paas et al., 2003b). Although three loads consume cognitive resources, they have distinct impacts on learning. Intrinsic and extraneous load tend to hinder learning, whereas germane load facilitates learning (Plass, Moreno, & Brünken, 2010). Specific effects associated with instructional designs have been provided through empirical research. For example, the split-attention effect (Chandler & Sweller, 1991) occurs when people must attend to multiple sources of information from the same presentation. In multimedia learning, adding on-screen text to animations or illustrations with simultaneous audio text results in a redundancy effect and leads to poorer performance (Kalyuga, Chandler, & Sweller, 1999). The redundancy effect is derived from a split-attention effect because multiple types of visual information compete for limited resources in the same channel and lead to distraction (Kalyuga et al., 1999).

Cognitive load theory of multimedia learning (Mayer & Moreno, 2003) is based on dual-coding theory (Clark & Paivio, 1991), the working memory model (Baddeley, 1992), cognitive load theory (Chandler & Sweller, 1991), generative theory (Wittrock, 1989), and the OSI model (Mayer, 1996). A series of experiments have been conducted to determine how different combinations of visual and auditory presentation can affect students’ learning. For example, Mayer and Moreno (1998) found evidence of split-attention effect using multimedia for teaching lightning formation. The results show that the group with animations and simultaneous narration outperformed the other groups. Moreno and Mayer (1999) manipulate the timing of narrations (including concurrent with, preceding, and following the animations) and confirm that words presented as audio text can facilitate learning better than words as written text, namely, the modality effect. Nevertheless, Moreno and Mayer (2002) find that, when no additional visual information (such as animations) appeared on the screen, students understood lightning formation better than when words were presented in both visual and auditory forms. Similar results are found in the study conducted by Chang et al. (2011). In this case, reading on-screen text with redundant narration helps students better comprehend scientific materials, provided that no animations interfere with words in the visual form. The results indicate that students can better integrate knowledge when information is presented in both the visual and the auditory channels and there is no concurrent interference in the same channel. Representations in both channels can be held in working memory simultaneously and be logically connected (Moreno & Mayer, 1999; 2002). In other words, adding on-screen text in addition to multimedia animations can overload the visual information-processing channel and lead to split-attention (Mayer, Lee, & Peebles, 2014).
Despite the advantages of dual-mode presentations, the effectiveness of presenting audio text and excluding written text or subtitles is dependent on the complexity and the difficulty of the text itself (Kalyuga et al., 1999). If the text (scientific knowledge involving unfamiliar terms and abstract concepts) is too difficult or too long to follow, then working memory will be overloaded. Therefore, the aforementioned multimedia design principles may not fit for learning materials that are unfamiliar to and difficult for learners, especially for EFL learners. Instead, EFL learners may better comprehend the narration with the presence of subtitles when viewing multimedia materials because reading skills are typically more developed than listening comprehension (Danan, 2004). The split-attention effect of subtitles has rarely been explored in EFL conditions. To fill this gap, the present study investigates the split-attention effect of viewing subtitles from the cognitive load perspective for EFL students learning scientific material through multimedia programs.

**Multimedia note-taking**

Note-taking is one of the most prevalent learning strategies in educational fields (Kauffman, Zhao, & Yang, 2011). Students take notes on papers when reading books or listening to lectures (Kauffman, 2004). Kauffman et al. (2011) concluded that encoding and external storage are the two important functions of note-taking. The encoding function aims to encode summarized information into long-term memory, whereas the storage function can keep self-generated material for subsequent review. Although notes serve as external storage to retain temporary information for elaboration and internalization, note-taking consumes cognitive resources, leading to high cognitive load especially during lectures (Piolat et al., 2005). Note takers need to encode the content by relating it to prior knowledge, which is similar to processing information by using an organizer to integrate multiple pieces of information into a coherent structure. According to cognitive load theory, such automatons and schema acquisitions can be beneficial to learning (Sweller, van Merriënboer, & Paas, 1998). Nevertheless, note-taking requires students to split their attention between listening to the instructor and jotting down their notes. Therefore, the effect of note-taking on learning outcomes may be positive (Moos, 2009), negative (Dunkel, 1988), or neutral (Faraco, Barbier, & Piolat, 2002) depending on students’ prior ability.

The proliferation of the use of multimedia in education provides learners a variety of learning options. Learning through multimedia in the form of videos has been shown to enhance learners’ motivation and performance in online courses (Griffin, Mitchell, & Thompson, 2009; Choi & Johnson, 2005). Several video note-taking tools were developed to fit in the dynamics of digital learning environment (see Rich & Hannafin, 2009 for a review).

When viewing videos, people can take notes on papers. However, without referencing the original video, the notes are subject to errors and not readily usable (Mu, 2010a). Mu (2010a) developed a video annotation system that allows users to directly link notes to a specific scene in the video. Comparing video annotations with and without direct links (i.e., Smartlink versus non-Smartlink) to the video content, Mu (2010a) finds that the Smartlink group took fewer notes and focused more on the video content rather than the video control (with evidence from eye-tracking) due to reducing the non-learning-related cognitive load.

**The present study**

We used an annotation system similar to Smartlink (Mu, 2010b) to explore the effect of subtitles and note-taking in the multimedia learning of science. We intend to answer the following questions.

- **Will reading subtitles result in a split-attention effect on the post-achievement, and cognitive load of EFL undergraduate students?** It is expected that reading subtitles will result in a split-attention effect, which in turn will impair learning performance and increase cognitive load.

- **Will note-taking on the multimedia material benefit or hinder the post-achievement, and cognitive load of EFL undergraduate students?** Given that note-taking may influence the extent to which the learners fixate on the ongoing video, it is hypothesized that note-taking hinders learning performance and increases cognitive load.

We employed a 2 (subtitles versus no subtitles) by 2 (enote-taking versus no enote-taking) factorial design and controlled for students’ English proficiency to investigate the effects of subtitles and enotes on students’ perceived cognitive load in listening, animation understanding, subtitle viewing, and enote-taking, in addition to post-performance. Post-performance is considered a measurable index with regard to cognitive load (Paas & van Merriënboer, 1994).
Methods

Participants

The participants were 73 non-biology major undergraduate students (52% female) recruited from two universities in northern Taiwan. All students were native speakers of Mandarin Chinese, with English as a foreign language. The students participated in the multimedia experiments in return for a small amount of credit toward their course grade. The participants were randomly allocated to one of the intervention groups. A total of 17 students were in the “subtitles only” group (ST), 15 were in the “enote-taking only” group (NT), 17 were in the “subtitles and enote-taking” group (ST-NT), and 18 were in the control group with no subtitles or enote-taking supports (Control).

The multimedia program: Brain anatomy and cognitive function

Brain anatomy is the focus of this experiment. Knowledge of brain anatomy is essential in scientific domains such as biology, neuropsychology, and medical education. Each area of the brain is linked to a specific function that has implications for learning, cognition, and physical health.

The program was constructed using Adobe CS5 Production Premium video production software. The four versions (ST, NT, ST-NT, or Control) of the video that were delivered to the four treatment groups had exactly the same animation and narration and differed only in the availability of English subtitles, which were presented concurrently with animation and narration, or the enote function. The subtitles functioned as an auxiliary tool to comprehend the meaning of the animation, whereas the enotes could help memory retention. All videos included a 15-minute animation depicting brain anatomy and the corresponding function areas of the brain (see Figure 1). The animation could be divided into two segments. The first part of the animation (i.e., the first 160 seconds) was a general introduction to the brain, including size, weight, and memory functions. The second part of the animation was a 3-D mapping of a 360° moving brain, with details regarding the functions of each area and segments that popped up for explanations and the introduction of neural transmitting routes. The materials included the essential information found in introductory science textbooks.

The enote tool

The web-based Anchor Instruction (WAI) developed by Wang (2010) is an enote tool designed for the effective annotation of video or e-instruction systems. The system interface consists of video playback buttons, a timeline control bar, annotation input boxes, and a historical annotation list, as shown in Figure 1.

![Figure 1. A screenshot of the brain animation with annotations](image)

(a) The playback screen and the control panel. (b) The textbox is used to input enotes. (c) The historical annotation list.

With the timeline control bar, students can play or pause the animation; moreover, they can set the play-head position by a mouse click to jump the video content to any frame, helping students to more efficiently extract the embedded information. In the annotation input area “b” in Figure 1, students can stop the video playback and add notes, such as questions, keywords, associations, and summarization. The annotations are immediately listed
on the right of the screen. If students click specific annotations on the historical list, the play-head position will accordingly move to where the annotation is anchored. AJAX (Asynchronous JavaScript and XML) technology is adopted to facilitate annotation in the streaming video in a non-interrupting manner. Proposed by (Garrett, 2005), AJAX is a newer data transmission technology that can update data without the re-arrangement of web pages.

The instruments

Pretests: English proficiency and prior knowledge

We included two pretests to evaluate students’ English proficiency and prior knowledge on brain anatomy. English proficiency was evaluated using the well-known General English Proficiency Test (GEPT), developed and validated by the Language Training and Testing Center in Taiwan. The sample test of reading comprehension at the high-intermediate level was used because college graduates are required to pass the high-intermediate level GEPT. The reading comprehension test consists of 25 multiple-choice items.

Ten multiple choice items were used to evaluate participants’ knowledge of the fundamental facts of brain components and cognitive functions; for instance, “What cognitive function is the hippocampus mainly in charge of?” Participants received one point for each correct answer and zero points otherwise. To ensure that students were listening to the narration, a manipulation check on the listening was performed with two multiple-choice items that require students to confirm the sentences that they heard in the program.

Posttests

The posttest consisted of two parts, a post-performance test and a cognitive load measure. The performance test included 14 multiple-choice items covering retention, image mapping, and inference on brain anatomy, with a total of 14 possible points. Learners’ post-achievements were measured when participants finished the task because achievement is considered a measurable index with regard to cognitive load (Paas & van Merriënboer, 1994).

In addition, students were required to answer questions regarding their mental effort spent on listening to the narration and watching the animation. The subtitles-only group needed to rate its mental efforts when reading subtitles. Similarly, the enote-taking-only group needed to rate its mental effort while taking notes. The subtitles and note-taking group needed to answer four questions on mental effort. The cognitive loads in performing the activities were measured on a 7-point Likert scale (e.g., “The subtitles in the video make me feel difficult.”) ranging from “very disagree” to “very much agree” (Park et al., 2011), which was modified based on research by Paas (1992). Higher score indicated higher mental effort (Kalyuga, Chandler, & Sweller, 2000).

Procedure

First, the participants were briefed on the procedures for the experiment and took the pretest. Afterward, the experimenter demonstrated how to use the enote tool. The time frame for the experiment was set to 70 minutes based on a previous test. Then, the participants worked through the multimedia program of the randomly assigned version (ST, NT, ST-NT, or control) at their own pace, which means that the participants could rewind the video to where they felt confused or skip frames that were trivial. During the experiment, two assistants monitored the participants and helped them when they had difficulties in proceeding or had questions when taking the tests. After learning with the multimedia program, the participants completed the cognitive load measures and the post-achievement test. When the participants finished all of the tests, they were given a goody bag containing snacks and a drink as a token of appreciation. We found that all of the participants could finish the experiment in 70 minutes.

Data analysis

A two-step regression analysis was conducted, with cognitive load and posttest as the dependent variables and prior knowledge of English proficiency and prior knowledge on the brain as the predictive variables in the first step and the experimental conditions as the between-subject factor in the second step to test the effects of
subtitles and note-taking on learning when the effects of prior knowledge were taken into account. The group mean differences on prior knowledge, English proficiency, cognitive load, and post-achievement were tested using analysis of variance (ANOVA) with the SAS 9.3 General Linear Modeling (GLM) procedure. Equations (1) and (2) represent the general form of the model.

\[
\begin{align*}
\text{Post-test} &= \beta_0 + \text{Prior-KnowEng} + \beta_1 \text{Prior-KnowBrain} + \beta_2 \text{Experi-Condition} + \epsilon_1 \\
\text{CL} &= \beta_3 + \beta_4 \text{Prior-KnowEng} + \beta_5 \text{Prior-KnowBrain} + \beta_6 \text{Experi-Condition} + \epsilon_2
\end{align*}
\]

(1) and (2)

Results

The effects of subtitles and eNote-taking on mental efforts (cognitive loads)

Table 1 presents the descriptive statistics on English proficiency, prior performance, cognitive loads, and post-achievement. First, the four groups showed no difference on either prior knowledge, \(F(3,69) = 2.13, p > .05\), or English proficiency, \(F(3,69) = .60, p > .05\). The results indicate that English proficiency and prior knowledge were controlled for, as expected. A series of ANOVA procedures were conducted with the four cognitive load ratings as dependent variables. Levene’s test suggests that the assumption of homogenous variance was not violated.

| Table 1. Means and standard deviations of pre-test and post-test variables for the four groups |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|
| Variables                       | ST-NT (n = 17)  | ST (n = 17)     | NT (n = 15)     | Control (n = 18) |
| English proficiency             | M               | SD              | M               | SD              |
| Prior performance               | 17.79           | 4.04            | 16.06           | 5.76            |
| Post performance                | 8.37            | 1.54            | 7.29            | 2.05            |
| Cognitive loads                 |                 |                 |                 |                 |
| CL, Listening                   | 3.00            | 1.67            | 3.18            | 1.59            |
| CL, Subtitle                    | 3.21            | 1.44            | 2.65            | 1.58            |
| CL, eNote-taking                | 2.53            | 1.22            | --              | 3.11            |
| CL, Animation                   | 2.21            | 1.08            | 2.06            | 0.83            |
|                                  |                 |                 |                 |                 |

Note. ST-NT= Subtitles and eNote-taking, ST=Subtitles only, NT= eNote-taking only, Control = neither subtitles nor eNote-taking. * = unavailable data.

Cognitive load in listening to narration

English proficiency significantly affected the cognitive load of listening to narration, \(F(1, 67) = 18.13, p < .001, \eta^2 = .21\), whereas prior knowledge on the brain did not significantly affect the mental load of listening, \(F(1, 67) = .24, p > .05\). After the two covariates were removed, the results suggest a strong main effect of subtitles, \(F(1, 67) = 34.21, p < .001, \eta^2 = .34\). EFL participants learning with subtitles (ST-NT group CL,listening = 3.00 and ST group CL,listening = 3.18) reported a considerably lower mental load than those learning without subtitles (NT group CL,listening = 4.63 and Control group CL,listening = 5.28). Adding subtitles to the animations may significantly reduce the cognitive load of EFL students in listening to English narrations. Our findings conflict with the split-attention effect, which indicates that presenting animations simultaneously with subtitles and narration will increase cognitive load (Mayer & Moreno, 1998). Instead, we found that adding subtitles to animations considerably reduces cognitive load. Furthermore, there was no significant effect from note-taking, \(F(1,67) = .71, p > .05\), and the interaction was not significant, \(F(1, 67) = 1.36, p > .05\).

Cognitive load in reading subtitles

Among the four groups, only two of them (ST-NT and ST groups) watched animations with English subtitles. The results indicate English proficiency significantly affected the perceived cognitive load in reading subtitles, \(F(1, 32) = 10.42, p < .01, \eta^2 = .25\), but that prior knowledge on the brain did not, \(F < 1\). Meanwhile, the learners who took enotes (ST-NT group CL,subtitle = 3.21) reported a marginally higher cognitive load while viewing subtitles than did those who did not (ST group CL,subtitle = 2.65), \(F(1, 32) = 3.82, p = .059, \eta^2 = .11\). In summary, taking enotes does not impose a significant cognitive load on EFL students while watching animations with subtitles.
Among the four EFL groups, only two of them (ST-NT and NT groups) were required to take enotes while learning multimedia contents. Therefore, we examined the differences between these two groups. The results indicate that English proficiency significantly affected the perceived cognitive load for taking enotes, $F(1, 32) = 8.54, p = .006, \eta^2 = .20$, but that prior knowledge on the brain did not, $F < 1$. The main effect of subtitles was not significant, $F(1, 32) = 0.65, p > .05$, meaning that those who read the subtitles (ST-NT group $\text{CL}_{\text{enote-taking}} = 2.53$) or not (NT group $\text{CL}_{\text{enote-taking}} = 3.11$) reported approximately the same level of mental load in taking enotes.

### Cognitive load in viewing animation

English proficiency significantly affected cognitive load while viewing the animation, $F(1, 67) = 4.45, p < .05, \eta^2 = .07$. By contrast, prior knowledge on the brain did not, $F(1, 67) = .48, p > .05$. After the covariate effects were controlled for, the results suggest a marginal significant effect of subtitles, $F(1, 67) = 3.65, p = .06, \eta^2 = .06$; however, enote-taking had no significant effect on reducing mental load, and there was no significant interaction between the two factors, $Fs < 1$. The participants in the subtitles groups (ST-NT group $\text{CL}_{\text{animation}} = 2.21$ and ST group $\text{CL}_{\text{animation}} = 2.06$) reported slightly less cognitive load in viewing the animations than those in the non-subtitles groups (NT group $\text{CL}_{\text{animation}} = 2.68$ and Control group $\text{CL}_{\text{animation}} = 2.61$); however, the difference is trivial.

### The effects of subtitles and enote-taking on post-performance

To test whether providing subtitles or taking enotes would influence EFL undergraduates’ performance of brain anatomy and cognitive functions, a series of ANCOVA procedures was conducted. Results showed that both covariates had significant impacts on performance test, for English proficiency, $F(1, 67) = 27.94, p < .001, \eta^2 = .29$ and for brain prior concepts, $F(1, 67) = 5.30, p < .05, \eta^2 = .07$. After controlling for the covariates, a significant difference in performance between the participants who read the subtitles (ST group post-performance = 11.16 and ST-NT group post-performance = 10.82) and those who did not (NT group post-performance = 9.95 and Control group post-performance = 9.78) was revealed, $F(1, 63) = 4.30, p < .05, \eta^2 = .06$. However, the results show no significant effect of note-taking, $F(1, 67) = .01$, or any significant interaction between subtitles and enote-taking, $F(1, 67) = .73$. Apparently, the EFL undergraduates learned more science concepts from animations with English subtitles but not from taking enotes.

### Discussion

The present study investigated the effects of subtitles and enote-taking on learning brain anatomy and cognitive functions in a multimedia learning environment. We introduced English proficiency and prior science knowledge as covariates. The results indicate that learners’ performance was influenced by prior knowledge and English proficiency. English proficiency was a strong covariate exercising effects on post-performance and the four types of cognitive load. EFL students with higher English proficiency learned better and perceived lower cognitive load; by contrast, students with higher prior knowledge outperformed in the post-test, but their higher prior knowledge had no effect on the ratings of the four cognitive loads. English listening and reading abilities seemed to be a critical precondition in learning scientific materials when the materials were presented in English to the EFL undergraduates.

After removing the covariates, the EFL undergraduates with the aids of subtitles had a better performance in the posttest and reported a lower cognitive load with regard to listening to the English narrations. The finding is consistent with the conclusion made by Mitterer and McQueen (2009), which claimed that foreign subtitles were beneficial in a movie learning context. Nevertheless, this effect on the cognitive load of viewing an animation was marginal. Our major finding suggests that, for EFL undergraduates, a multimedia program accompanied by both English narration and on-screen English text is less likely to result in a split-attention effect even through such an effect has been proven in native learners (Mayer & Moreno, 2003). The split-attention effect (shown on the left of Figure 2), caused by adding subtitles beside multimedia animations, can overload native English speakers’ visual information-processing channel and lead to split-attention between two visual sources (Moreno & Mayer, 1999; Tabbers, Martens, & van Merriënboer, 2004). By contrast, our EFL participants learned better with the aid of subtitles (shown on the right in Figure 2). Furthermore, listening comprehension of English
narration accompanied with subtitles imposed less cognitive load on learners than those without subtitles did. Our comparable finding with the previous research is discussed in the following, with Figure 2 as a visual aid.

![Figure 2. A comparison between native learners and EFL learners with regard to the effect of subtitles](image)

Based on participants’ reports, the major cognitive load seemed to stem from English listening comprehension. When compared with viewing the animation, reading subtitles and taking enotes, learners reported listening comprehension as a major source of cognitive load. Listening comprehension is regarded as an active process in which an individual concentrates on selected aspects of audio input, forms meaning from passages, and associates what was heard with existing schema (O’Malley, Chamot, & Kupper, 1989). In particular, listening comprehension requires considerable cognitive resources, and for most EFL learners, it is processed in an effortful manner (Kurita, 2012). This finding can be explained by the research of Underwood, Kenworthy, and Rost (1989), in which possible obstacles to listening comprehension are noted. First, listeners cannot control the speed of delivery. If the conversation is too fast, then EFL listeners are likely to ignore/miss some critical words. In listening comprehension, even the shortest interruption in attention can severely impair comprehension. Second, EFL listeners may have a limited vocabulary. The speaker may choose words that the listener does not know. Listeners sometimes encounter unfamiliar words, causing them to stop and retrieve the meanings of the words from memory, which can cause learners to skip the next part of the speech. Third, the lack of contextual knowledge increases the difficulties in sharing mutual knowledge and common content.

Additionally, the visual channel is required to process tremendous volume of animation (2D or 3D brain anatomy - sometimes rotated 360° – and other animation metaphors) and on-screen subtitles simultaneously (left hand side in Figure 2). Meanwhile, the audio channel is required to process English narration, which presented redundant information as the subtitles. In the current study, our EFL students did not process English narration with ease so that they might be inclined to allocate most of their attention to view the subtitles for comprehension. This finding is in accordance with the suggestion of previous studies (e.g., Kurita, 2012), which claim that reading is relatively easy and automatic for EFL students whereas listening comprehension is not. For EFL students with relatively high listening comprehension skills, the narration may be treated as auxiliary information that is only for mapping or in aid of the comprehension of the subtitles. For those with low listening skills, understanding may be primarily dependent on viewing the subtitles. We believe that, without subtitles, basic listening comprehension is barely achieved. This implies that for EFL students the narration processed in the audio channel could be comprehended only when the subtitles in visual channel were comprehended; whereas, when the subtitles were not provided, the unfamiliar vocabulary or high speed of the narration may impair comprehension severely or totally.

Regarding the redundancy effect (the green arrow shown on the left in Figure 2), the results suggest that, although simultaneously presenting animations and subtitles may consume cognitive resources, evidence from the post-achievement test indicate that EFL learners learn better when they are presented animations with subtitles. Therefore, a reversed redundancy effect might exist for EFL learners. In other words, the subtitle may
impose cognitive load on EFL students, however, the subtitle might play as an auxiliary role for EFL students to learn better as well.

Our second finding suggests that taking e-notes while learning biology through animations is not beneficial for EFL learners who are non-biology majors. This finding is inconsistent with the previous findings (Makany et al., 2009), in which note-taking is beneficial for native adult learners. In general, we find that e-note-taking has no positive effects on learners in either increasing the post-test scores or reducing cognitive load. The cognitive load in note-taking was significantly influenced by learners’ English proficiency. Learners with better English proficiency reported a lower cognitive load while taking e-notes. In addition, no evidence suggests that subtitles had an impact on cognitive load in taking e-notes. We considered that watching animation could be an obstacle for learners to take e-notes. The learners in the note-taking group could pause at any time to note down whatever they considered important. However, no significant improvement on the post-achievement was observed. Furthermore, e-note-taking seems to be unable to reduce cognitive load in listening. This result can be interpreted by the SOI model (Mayer, 1999), which suggests that knowledge can be obtained by selecting, organizing, and integrating relevant information. Our results reflect the fact that, for EFL students, the major problem in learning without on-screen text is that learners may be unable to correctly comprehend the main gist of the narration. When EFL learners could not correctly interpret the acoustic information before the information slipped out of working memory, poor understanding was the result. Although taking e-notes is a cognitive aid in organizing and integrating concepts, the impairment of listening comprehension interferes with the effect of e-note-taking by overloading working memory or poor comprehension of the learning materials. Therefore, only limited knowledge could be acquired.

Finally, the present study should be interpreted with caution. In this study, prior knowledge is viewed as a covariate variable, not as a factor. Therefore, it is inappropriate to test the interaction among prior knowledge, subtitles, and e-note-taking. Future studies can employ prior knowledge as a factor to investigate the interaction among them. Another limitation involves validation with respect to subtitles. Although two items were designed to evaluate whether learners focus their attention on the subtitles, the criteria must be improved for better validation. Currently, it is crucial to investigate on-processing behaviors (e.g., eye movement) to increase our understanding of learning processes. Future research may highlight on-processing measure to strengthen the connection of the hypothesis to the inference. Moreover, Hubbard (2005) suggests that researchers can reduce bias by avoiding several characteristics in designing experiments. These characteristics involve a small sample size, objective evaluative surveys or questionnaires. In addition, participant characters that should be avoided include, engagement with the learning task for only a short time, untrained participants, no guidance during the study, inexperience with the task, and learning the material during their initial experience with the task.

Conclusion

To address EFL learners’ difficulties in learning multimedia materials, researchers analyze the crucial factors that can affect learners’ performance. In the future, we expect that more elaborate research can be conducted to gain understanding on these factors. For example, an eye tracker can be implemented to monitor learners’ processes and observe whether EFL learners focus on the subtitles while listening to the narration. By doing so we could have a clearer view as to what types of EFL learners ignore narration and merely use subtitles to learn. In addition, a within-subject design with multiple implementations is suggested to yield generalizable findings. Such a design can be beneficial in allowing each participant to experience different treatments and increase innovative learning opportunities. Moreover, the within-subject design can increase the statistical power and reduce error variance due to individual differences.

References


Comparing the Effectiveness of Self-Learning Java Workshops with Traditional Classrooms

Kiran L. N. Eranki* and Kannan M. Moudgalya
IDP Educational Technology, IIT Bombay, Powai, India // erankikiran@iitb.ac.in // kannan@iitb.ac.in
*Corresponding author

(Submitted August 9, 2015; Revised February 3, 2016; Accepted March 29, 2016)

ABSTRACT
In this work, we study the effectiveness of a method called Spoken Tutorial, which is a candidate technique for self-learning. The performance of college students who self-learned Java through the Spoken Tutorial method is found to be better than that of conventional learners. Although the method evaluated in this work helps both genders, females seem to benefit more through self-learning workshop based approach. Self-learners found to be more interested in ease of learning and the availability of content, while classroom learners focused more on evaluation. All of these findings are in agreement with the literature. Although the proposed method seem to be effective to a restricted class of topics, the benefits can still be enormous and suitable for flipped classrooms, MOOC, etc.

Keywords
Programming, Spoken Tutorials, Self-learning, Gender differences, Perception Analysis

Introduction
This study looks at the effectiveness of the Spoken Tutorial methodology, developed to provide IT training through the self-learning route (Moudgalya, 2011). Although a large use case of it is reported by Moudgalya (2014) with feedback from 25,000 people, no controlled study of it has been carried out so far. As this methodology has the potential to provide computer literacy to a large number of people and thus improve their employment potential, it will be beneficial to understand its strengths and ways to improve it.

According to Moudgalya (2014), self-learning capability is the main reason for the widespread acceptance of Spoken Tutorials. Self-learning seeks to put the learner as much as possible in control of the learning process. This mode of learning involves behavioral motivation and meta-cognitive understanding of social perspectives through self-learning (Pintrich, 1999). Self-learners proactively seek out information when needed and prepare to master them. Self-learners use obstacles as opportunities to succeed (Zimmerman & Martinez-Pons, 1990). Students’ perception of self-learning shows both a motive to learn and subsequent outcomes of effort to succeed (Schunk et al., 2012). Self-learners select learning strategies to achieve desired outcomes based on the feedback received from learning outcomes and levels of difficulty (Hiltz & Wellman, 1997). Socio-cognitive learning models proposed by Zimmerman (1989), Boekaerts et al., (2005) have shown that learners performed better through Self-learning approach. On the other hand, studies on classroom show that students fail to interact with teachers to clarify their doubts despite the opportunity due to low self-esteem, which may be because of instructor/parent-dependence and poor self-control (Schmid et al., 2014). Othman et al. (2014) have found that the teaching method based on self-learning enhanced students’ skill of using application software in computing environments. Moreover, based on these results, they suggest that instructors should help students regulate their learning by applying self-learning instructional methods in e-learning environments.

Studies in computing education have shown differences among male and female student perceptions and learning outcomes: males seem to find programming easier, showing higher intentions and higher learning outcomes than female students (Guzdial et al., 2014; Alvarado et al., 2014; Carter & Jenkins, 1999). Gender differences have also been noticed among Asian and Indian learners across all social and socio-economic groups with serious issues on parental control and education delivery systems (Arnold, 2015). The above research studies imply that learners benefit from self-learning approach, by engaging them into planning, monitoring, adaptability and evaluation phases of learning.

Almost all the studies on self-learning have been carried out in a classroom setting. In the Indian context, it is important to make the self-learning feature available through workshops also. To the best of our knowledge, workshop based self-learning approach has not been studied for its effectiveness. In the current study, we use the Spoken Tutorial approach to train learners using self-learning workshops (Moudgalya, 2014).
Although a lot of papers have appeared on the Spoken Tutorial method, its effectiveness is not established through a rigorous study, which is one of the main contributions of this work. In addition to this, we also have identified self-learning characteristics of workshop and classroom learners through this study. We also study the effect of gender on self-learning.

**Spoken Tutorial methodology**

Spoken Tutorial project is an initiative of National Mission on education through Information and Communication Technology (NME-ICT), Ministry of Human Resource Development (MHRD), Govt of India (Sakshat, see http://www.sakshat.ac.in/). Spoken Tutorials are developed to improve the quality of programming education in India. Each Spoken Tutorial consists of a 10-12 min screencast of an actual session, demonstrated by an expert, illustrating various programming concepts, as shown in Figure 1. This recording is accompanied by the narration with a script, As Spoken Tutorials are created for self-learning, and it is possible to provide large scale training, thereby addressing the shortage of employable youth in India (Ray, 2014). All the content developed through this project is attributed under a creative commons license (CC BY SA).

Spoken Tutorials are made suitable for self-learning through novice check of scripts before creating the tutorial. The self-learning feature is verified through pilot workshops. These steps are similar to the ones proposed by Moseley et al. (2005). Spoken Tutorials follow the example-centric and example-based approach suggested by Neal (1989). Spoken Tutorials attempt to reduce the cognitive overload by (i) juxtaposing audio and video (ii) keeping sentences short (iii) taking time to explain difficult or new concepts and (iv) recording an actual session (Moudgalya, 2014). Cognitive overload and temporal contiguity are two impediments for self-learning (Moreno & Mayer, 1999). Localization and the use of short sentences make Spoken Tutorials accessible to students who are not fluent in English (Barac & Bialystok, 2012).

Spoken Tutorial based Education and Learning through Free and Open Source Software (FOSS) study workshops are abbreviated as SELF (Spoken Tutorial based Education and Learning on FOSS) workshops. SELF workshops are conducted to help students with limited access to computers, bandwidth and access to Internet, during college hours (Moudgalya, 2014). Organizing workshops through colleges provide the following benefits: it becomes an accepted academic activity; all systems can be pre-loaded with Spoken Tutorials and required software. SELF workshops allow students to learn at their own pace, using a language of their choice, as the audio of Spoken Tutorials is dubbed into all 22 languages of the Eighth Schedule of the Indian Constitution.
We will briefly point out the reasons why the Spoken Tutorial approach provides large scale training. The ability to download an entire collection of Spoken Tutorials for offline use obviates the need for Internet access in every computer system, one of the biggest bottlenecks for ICT based learning in India. SELF workshops are conducted by volunteers, who need not be experts. A volunteer who conducts a workshop on a topic can easily conduct several others as well. As no software is to be bought and as SELF workshops are offered free of cost, administrators can easily approve this training. Finally, as students trained through this method do well in exams and also get better jobs, many educational systems are including Spoken Tutorials as a part of their curriculum (Tamil Nadu Department of Technical Education, 2013).

We will now discuss the pedagogical benefits of Spoken Tutorials. The side-by-side method encourages a student to reproduce every action that is demonstrated in a Spoken Tutorial, resulting in active learning see Figure 2. The benefits of active vs. passive learning are well known (Grissom et al., 2003; Kroll & Laboskey, 1996). Learning at a convenient pace allows students to resolve their doubts then and there. Lack of this opportunity in synchronous lectures often leads to doubts not getting cleared at all (Vrasidas, 2000). Localization of tutorials allows students to learn new and complicated concepts also easily (Alanís & Rodríguez, 2008). As every Spoken Tutorial comes with an assignment, a student is forced to try out their understanding immediately, a facility that may not be available in conventional methods (Hung et al., 2010). Students can use Spoken Tutorials as a reference material and to practice difficult topics even after the completion of SELF workshops (Eranki & Moudgalya, 2013). SELF workshops help scale up training through Spoken Tutorials. Raval et al. (2012) show that workshop based approach to teach students and teachers is effective.

**Research questions**

Not much research is available on self-learning workshops in the context of non-conventional education practices in India. As explained above, there are many benefits if the results of self-learning are comparable with that of the classroom methodology, the main one being large scale training without quality dilution. Moudgalya (2014) has demonstrated that Spoken Tutorial based training is scalable. But the effectiveness of this method is not demonstrated with a controlled, rigorous, study. So, we arrive at the first research question.

- Are Spoken Tutorial based self-learning workshops as effective as traditional classrooms?

If the answer to the above question is positive, and if we can characterize self-learners, it will help us determine to which type of students the self-learning methodology proposed in this study will be effective. If the perception of students matches the actual performance, it will increase the confidence level of our approach.
So, we arrive at the next research question:

- **What are the characteristics of self-learners and are their perceptions in agreement with actual performance?**

Females generally tend to perform worse than males in programming (Lee, 2014). As a result, here is a possibility of this happening in the proposed approach as well. As a matter of fact, there is a possibility of females performing much worse than males. If this were to happen, the proposed method would have to be rejected as ineffective. So, we arrive at the final research question:

- **Does the gender difference influence learning?**

**Methodology**

In this section, we will describe about the research methodology applied to evaluate the student perceptions and validate them with actual performance in Java post-test. We begin with sample for the study, followed by Self Regulated Learning (SRL) questionnaire used to predict self-learning ability of the participants and further, validation of student perceptions with actual Java test.

**Java SELF workshops and classroom course, compared with post-test**

We will first explain the control and experimental groups we selected for this study. A total of 400 students were randomly selected for the study and distributed among the control and experimental groups. We selected these students as most of them have expressed having watched Youtube (http://www.youtube.com) online videos outside the classrooms or at home, either to complete the coursework or seek better understanding of concepts based on pre-workshop feedback.

The control group consisted of 180 students, studying at a local engineering college. The experimental group consisted of 220 students from the same college. We established the equivalence amongst these groups using a pre-workshop questionnaire. None of these 400 students had exposure to any programming language. All of these students were of same age group (below 25), from different departments, other than Computer Science. The participants from the experimental group (n = 220) attended a three hour Spoken Tutorial based SELF workshops on Java programming course. As a SELF workshop could accommodate a maximum of 40 students, owing to the limited computer infrastructure, the experimental group underwent Java training in six batches. The control group (n = 180) attended conventional Java classroom lectures, conducted by a teacher on topics similar to the SELF workshop. They attended five lectures of one hour duration each.

Next, we will explain how the course was conducted for the two groups. We choose the following ten Java concepts for the test: Operators, Arrays, Constructs, Classes, Methods, Inheritance, Polymorphism, Overloading, Constructors, Modifiers. The classroom students learnt these topics in five one hour lectures, learning two topics per lecture. On the other hand, the workshop students studied the same ten topics in a three hour SELF workshop. The post test included questions of three different levels of difficulty (apply, analyze, evaluate) based on revised blooms taxonomy (Thompson et al., 2008).

**Zhao and Chen’s SRL questionnaire**

We have used Self Regulated Learning questionnaire (SRL) to study the implications of self-learning among the workshop and classroom learners. Background work by Zimmerman (1989); Schunk et al. (2012); Pintrich (1999) included:

- Validity and reliability of SRL Questionnaire was already established through studies conducted by Zimmerman and Martinez-Pons (1990); Zimmerman (1989); Pintrich (1999).
- According to Zimmerman (1989), Zimmerman, Bandura and Martinez-Pons (1992) and Schunk et al. (2012), self-learning instructional content should contain four ingredients: Self-evaluation and monitoring; Strategy to plan learning outcomes; Adaptability; and Strategy to monitor learning outcomes.
Zhao et al. (2014) found the same four to be applicable to other types of self-learning situations, such as online and distance learning modes. They identified three secondary characteristics for each of the above mentioned four primary characteristics. These are (i) Learning objectives (LO), (ii) Learning resources (LR) and (iii) Learning outcomes (LC).

- **Learning Objective [LO]** focuses on the interest of a learner towards conceptual knowledge, programming skill and motivation.
- **Learning Resource [LR]** refers to course learning materials, method of teaching-learning for supporting and improving programming skills.
- **Learning Outcomes [LC]** focuses on conceptual understanding of basic programming skills, mainly comprehension and debugging skills of the participants.

Zhao et al. (2014) validated a questionnaire of 57 questions on distance learning, as shown in Figure 3. The breakup of these 57 questions is given in Table 1.

### Table 1. Questionnaire of Zhao and Chan: Questions in primary and secondary characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>LO</th>
<th>LR</th>
<th>LC</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>Monitoring</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td>Evaluation</td>
<td>6</td>
<td>4</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>Evaluation</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>21</td>
<td>17</td>
<td>19</td>
<td>57</td>
</tr>
</tbody>
</table>

Zhao et al. (2014) validated a questionnaire of 57 questions on distance learning, as shown in Figure 3. The breakup of these 57 questions is given in Table 1.

### Figure 3. Self-Regulated Learning Questionnaire secondary characteristics (Zhao et al., 2014)

**Applicability of Zhao and Chen’s questionnaire in our study**

This questionnaire is applicable also to workshop based, self-paced, learning process, because in the Spoken Tutorial method also, learning is individualistic and not group based. Spoken Tutorials, created for self-learning, play the role of mentors. This claim is validated by comparing the results of questionnaire with the actual performance in a test. We only changed the phrase *distance learning* by the word workshop in this study. Student
perceptions were captured through this questionnaire to evaluate their self-learning behavior and motivation to learn Java programming course.

Sample for Zhao and Chen’s SRL questionnaire

- The questionnaire was administered to 420 participants who gave their consent to participate in the study.
- A total of 400 questionnaires were received (response rate of 98.6%) of which 20 questionnaires were eliminated for incomplete submissions.
- Students who made up the control group for the Java post-test formed the control group for this study also. The experimental group is also identical in the same way. All participants voluntarily participated in this study and gave their consent to participate.

Results and discussions

In this section, we will present answers to the three research questions raised earlier. We will first compare the effectiveness of the Spoken Tutorial methodology in self-learning. We will next characterize self-learners. We will conclude this section by answering our question on gender differences.

RQ1: Effectiveness of Spoken Tutorial methodology

We have several reasons to believe that the workshop participants have learnt Java at least as well as the classroom students, thereby answering the first research question:

- The workshop learners had an average score of 69.09%, while classroom learners scored 64.8%. It is also statistically significant ($t = 6.03, p < .01$).

- The Cohen $d^{X1−X2}$ effect size was 0.623, among workshop and classroom groups, which indicates a significant effect of treatment in improving the programming skill through workshop. As it is greater than 0.6, this values indicates a large effect size among the groups.

- The students were further divided into three groups, based on their scores: high ($\geq 80$), medium ($\geq 60$, but $< 80$) and low ($< 60$). The numbers of students who fall in these three groups are shown in Figure 3. We found 13% of the students of the workshop fall in high scorers group, 57% as medium scores, and 30% as low scorers. The corresponding numbers in the classroom group are 5%, 50% and 45%. The number of high scorers in the workshop group are two and a half times the corresponding number in the classroom group. The workshop group has done better in the medium score category also.

- We found 57% of workshop students fall in medium scorers group, 30% of them as low scorers and 13% of them fall in high scorers group. While classroom group had almost equivalent distribution with 45% of them among the low and 50% of them in medium groups. And only 5% of them fall in high scorers group, which is lesser than workshop group, see Figure 4.

- The average marks scored by the workshop students on apply, analyze and evaluate type of Java questions are 72%, 71% and 65%, respectively, while the corresponding classroom scores are 66%, 60% and 62%.

- It is only natural that the workshop group has done better than the classroom group overall. As a matter of fact, in every one of the three of apply, analyze and evaluate categories of Sec. 4, the workshop team has done better, as can be seen in Table 2.

- Although in general males perform better than females in computer programming as found by Young (2014), it is interesting to note that the workshop females have actually done better than classroom males, see Table 3. This will be taken up for a detailed discussion while answering RQ-3. Naturally, they have done better than classroom females as well. Of course, they perform worse than workshop males, which is consistent with (Cassidy & Eachus, 2002).
Figure 4. Division of workshop and classroom participants into low, medium and high groups, based on the marks they scored.

Table 2. Marks scored by workshop and classroom groups in apply, analyze and evaluate categories

<table>
<thead>
<tr>
<th>Category</th>
<th>Workshop (%)</th>
<th>Classroom (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apply</td>
<td>72</td>
<td>66</td>
</tr>
<tr>
<td>Analyze</td>
<td>71</td>
<td>60</td>
</tr>
<tr>
<td>Evaluate</td>
<td>65</td>
<td>62</td>
</tr>
</tbody>
</table>

Table 3. Java test, inter-group and intra-group, analyses

<table>
<thead>
<tr>
<th>Groups</th>
<th>Gender (participants)</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Workshop</td>
<td>Overall_males</td>
<td>67.5</td>
<td>1.55</td>
<td>2.58</td>
<td>.009**</td>
</tr>
<tr>
<td></td>
<td>Overall_females</td>
<td>65.6</td>
<td>1.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Workshop_male</td>
<td>69.9</td>
<td>1.32</td>
<td>2.31</td>
<td>.02**</td>
</tr>
<tr>
<td></td>
<td>Workshop_female</td>
<td>67.45</td>
<td>1.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Classroom_male</td>
<td>64.7</td>
<td>1.66</td>
<td>0.603</td>
<td>.546</td>
</tr>
<tr>
<td></td>
<td>Classroom_female</td>
<td>63.95</td>
<td>1.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Workshop_male</td>
<td>69.9</td>
<td>1.32</td>
<td>5.47</td>
<td>.001**</td>
</tr>
<tr>
<td></td>
<td>Classroom_male</td>
<td>64.7</td>
<td>1.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Workshop_females</td>
<td>67.45</td>
<td>1.76</td>
<td>2.05</td>
<td>.04**</td>
</tr>
<tr>
<td></td>
<td>Classroom_male</td>
<td>64.7</td>
<td>1.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Workshop_male</td>
<td>69.9</td>
<td>1.32</td>
<td>6.03</td>
<td>.0066*</td>
</tr>
<tr>
<td></td>
<td>Classroom_female</td>
<td>63.95</td>
<td>1.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Workshop_females</td>
<td>67.45</td>
<td>1.76</td>
<td>2.54</td>
<td>.01**</td>
</tr>
<tr>
<td></td>
<td>Classroom_females</td>
<td>63.95</td>
<td>1.54</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

- The perception of the workshop group was also better than that of the classroom group. Workshop male learners show statistically significant score for adaptability-learning objective ($t = 2.83, p < .01$) when compared with classroom males. Similarly, workshop females also show significance for adaptability-learning objective ($t = 1.97, p < .05$) when compared with classroom females. In the era of lifelong learning, traditional learning can no longer satisfy all learning needs, for the following reasons: (1) High quality teaching requires sufficient number of suitable teaching staff; (2) Studies in public universities/institutes are expensive; thus, accessibility is usually limited and subject to budget cuts and restrictions; (3) Traditional learning is restricted to a particular place, specific time, and a uniform pace (Beller & Or, 1998). Learners who opt for self-learning usually adapt themselves to gather more information before they plan to pursue a course and this also involves informed choice and commitment to master the course owing to the above mentioned challenges. We found similar aspirations in the workshop learners as compared to classroom learners.

RQ2: Traits of self-learners and correlation with performance

We will begin with the perception of learners. We computed the SRL questionnaire scores for all the 57 questions of Zhao et al. (2014) given in Table 1. The average score obtained for all students in both workshop
and classroom groups is 4.00. As this is above 3.8 on the 5 point Likert scale, we can conclude that all the students, irrespective of groups, are serious about the Java course. One possible reason for this is the employment potential. We consider null hypothesis $H_0$ as no difference among the workshop and classroom learners for all four characteristics of the SRL questionnaire. We performed a $t$-test on questionnaire responses to identify the differences between workshop and classroom learners. Although the average score of the workshop participants for all the 57 questions of Table 1 was higher than that of classroom students (Workshop = 4.06, Classroom = 3.94), and statistically significant ($t = 2.56$, $p < 0.01$), see Table 6. Naturally, there are statistically significant differences in individual characteristics.}

**Figure 5.** Scores of workshop and classroom participants in first level characteristics of the SRL questionnaire

Workshop learners scored higher than the classroom learners in all four first level characteristics, see Figure 5. They are also statistically significant in planning ($t = 3.35$, $p < .01$) and monitoring characteristics ($t = 2.47$, $p < .05$), see Table 4.

**Table 4.** SRL questionnaire first-level characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Group</th>
<th>Second-Level</th>
<th>Mean</th>
<th>SD</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning</td>
<td>Workshops</td>
<td>PLA-LO</td>
<td>3.98</td>
<td>0.7998</td>
<td>1.93</td>
<td>.0053**</td>
</tr>
<tr>
<td></td>
<td>Classroom</td>
<td>PLA-LO</td>
<td>3.83</td>
<td>0.7621</td>
<td>1.93</td>
<td>.0505**</td>
</tr>
<tr>
<td></td>
<td>Workshops</td>
<td>PLA-LR</td>
<td>4.12</td>
<td>0.7015</td>
<td>3.46</td>
<td>.0056**</td>
</tr>
<tr>
<td></td>
<td>Classroom</td>
<td>PLA-LR</td>
<td>3.85</td>
<td>0.7435</td>
<td>3.46</td>
<td>.0056**</td>
</tr>
<tr>
<td></td>
<td>Workshops</td>
<td>PLA-LC</td>
<td>4.46</td>
<td>0.3684</td>
<td>7.54</td>
<td>.0016**</td>
</tr>
<tr>
<td></td>
<td>Classroom</td>
<td>PLA-LC</td>
<td>4.05</td>
<td>0.6502</td>
<td>7.54</td>
<td>.0016**</td>
</tr>
<tr>
<td>Monitoring</td>
<td>Workshops</td>
<td>MNT-LO</td>
<td>3.98</td>
<td>0.5009</td>
<td>0.195</td>
<td>.855</td>
</tr>
<tr>
<td></td>
<td>Classroom</td>
<td>MNT-LO</td>
<td>3.96</td>
<td>0.5147</td>
<td>0.195</td>
<td>.855</td>
</tr>
</tbody>
</table>

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

Results also indicate that workshop learners rated higher score for planning-learning objective, planning-learning resources and planning-learning outcomes than the learners from classroom ($t = 1.93$, $p < .05$; $t = 3.45$, $p < .01$; $t = 7.54$, $p < .01$), see Table 5. This confirms that the workshop group is better in every aspect of planning, as compared to the classroom learners. Monitoring-learning resources and monitoring-learning outcomes of workshop learners are also statistically significant as compared to classroom group ($t = 3.39$, $p < .01$; $t = 4.027$, $p < .01$). We also found statistical significance for adaptability-learning resources for workshop learners ($t = 1.93$, $p < .05$). This can be attributed to the difficulties of classroom learners to clarify their doubts and availability of resources (Dehnadi et al., 2009). Although workshop and classroom group scores do not differ significantly in adaptability and evaluation characteristics, we reject the $H_0$ null hypothesis as we have found statistically significant differences in planning and monitoring characteristics.

**Table 5.** SRL questionnaire second-level characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Second-Level</th>
<th>Group</th>
<th>Mean</th>
<th>SD</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning</td>
<td>PLA-LO</td>
<td>Workshop</td>
<td>3.98</td>
<td>0.7998</td>
<td>1.93</td>
<td>.0053**</td>
</tr>
<tr>
<td></td>
<td>Classroom</td>
<td>Workshop</td>
<td>3.83</td>
<td>0.7621</td>
<td>1.93</td>
<td>.0505**</td>
</tr>
<tr>
<td></td>
<td>PLA-LR</td>
<td>Workshop</td>
<td>4.12</td>
<td>0.7015</td>
<td>3.46</td>
<td>.0056**</td>
</tr>
<tr>
<td></td>
<td>Classroom</td>
<td>Workshop</td>
<td>3.85</td>
<td>0.7435</td>
<td>3.46</td>
<td>.0056**</td>
</tr>
<tr>
<td></td>
<td>PLA-LC</td>
<td>Workshop</td>
<td>4.46</td>
<td>0.3684</td>
<td>7.54</td>
<td>.0016**</td>
</tr>
<tr>
<td></td>
<td>Classroom</td>
<td>Workshop</td>
<td>4.05</td>
<td>0.6502</td>
<td>7.54</td>
<td>.0016**</td>
</tr>
<tr>
<td>Monitoring</td>
<td>MNT-LO</td>
<td>Workshop</td>
<td>3.98</td>
<td>0.5009</td>
<td>0.195</td>
<td>.855</td>
</tr>
<tr>
<td></td>
<td>Classroom</td>
<td>Workshop</td>
<td>3.96</td>
<td>0.5147</td>
<td>0.195</td>
<td>.855</td>
</tr>
</tbody>
</table>
We will now characterize self-learning.

- We have found high scores for planning-learning outcomes (4.46), monitoring-learning outcomes (4.38) and followed by adaptability-learning outcomes (4.02), evaluation-learning outcomes (3.96) for second level SRL characteristics, see Figure 6.
- These scores show that planning-learning outcomes was considered to be more important for self-learning and while evaluation-learning outcomes was least preferred, which is true, as most self-learners consider learning outcomes based on ease of learning, availability of content as first priority, followed by evaluation of learning especially in online and distance self-learning (Lee, 2014).
- We have found statistically significant difference among the workshop and classroom groups on adaptability learning objective and monitoring learning resources and monitoring-learning outcomes, see Table 5.
- These results are similar to the reciprocal relationship theory proposed by Schunk et al. (2012), when students set intermediate goals within their proximity of time, they tend to perceive focused learning and this reciprocally prepares them to set more challenging goals.
- In conclusion, the majority of self-learners have good planning, monitoring, adaptability and evaluating capabilities in relation to learning resources. This answers the second research question: the above said four first level characteristics seem to indicate a higher level of learning potential amongst the workshop group.
- Through a feedback that we collected separately, 78% of the workshop learners expressed an interest to learn other topics through SELF workshops. This shows that the self-learners we worked with, liked the Spoken Tutorial approach of self-learning.
These results are consistent with the previous research studies (Young, 2014; Hiltz & Wellman, 1997; Winne, 2010). These findings are closely related to the culture of learning in India and several other Asian countries, as explained next. University distance education programs provide provision for part-time learners to complete their courses through study-centers. Students studying through this mode of classroom-independent teaching-learning also showed similar perceptions to self-learning (Iyer, 2014). Educational reforms allowed part-time study for employed adult learners and full-time study opportunities for high school graduates. These options facilitate independent study with less intervention through tele-learning centers and online portals. However, most of these programmes failed to meet the requirements of distant learners due to lack of bandwidth, infrastructure and content delivery issues. In the current approach, bandwidth and the infrastructure are not an issue, as instructional material and the target software were installed in every computer locally.

RQ3: Gender differences

We conducted a full multivariate analysis of the data to determine, if any gender differences are present in the sample. We conducted the t-test to study the gender differences among the groups. We first present the Java posttest comparison, before discussing perception.

Overall females vs overall males

- We found statistically significant differences among the overall male and female learners ($t = 2.58, p < .01$), see Table 3. Overall males scored 67.5%, while overall female score was 65.6%, confirming males scored higher than females among both the groups. These results can be attributed to the gender differences among programming skills, computer usage, ownership and access issues which is a major challenge among the female learners (Calero et al., 2007).

- Perception scores also showed significant gender difference across the data as a whole ($t = 2.72, p < .05$), suggesting that males are significantly better across all four first level characteristics than females. We found statistically significant differences among the overall male and female learners for planning ($t = 4.79, p < .01^*$) and monitoring ($t = 2.22, p < .05^*$). However no such significance was found among adaptability and evaluation characteristics. The gender differences among the groups for all four characteristics has been shown in Tables 6-8. Gender differences lead to poor social presence and disparities of opportunity among men and women. After doing a detailed study of Chinese and Indian Women, Harish (2014) points out that the only way to address this issue is to train women on self-learning methods. Because of sociological and historical reasons, men are more encouraged to take up employment compared to women. We have shown below that the SELF workshops help workshop women considerably- they end up doing better than classroom males as well. The fact that they can do this despite their initial handicap is because of the self-learning nature of the SELF workshops. Thus, we can say that the self-learning helps women even more than it helps men.
Table 6. Gender perception analysis -I

<table>
<thead>
<tr>
<th>Group</th>
<th>LO</th>
<th>LR</th>
<th>LC</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Workshop</td>
<td>4.06</td>
<td>2.56</td>
<td>.010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classroom</td>
<td>3.94</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Workshop_males  | 4.012| 2.72| .0068*
| Workshop_females| 4.125|     |     |      |      |
| Class_males     | 3.98| 0.9534| .3416|
| Class_females   | 3.93|     |     |      |      |

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

Workshop males vs workshop females

- Workshop males showed higher performance in Java test as compared to workshop females. Java test scores were also statistically significant among the workshop males and females \((t = 2.31, p < .05)\), see Table 3.
- Perception scores also showed statistically significant differences among the workshop male and female students for planning-learning objective \((t = 0.0019, p ≤ .01)\), planning-learning resources \((t = 0.0039, p ≤ .01)\), planning-learning outcomes \((t = 0.0010, p ≤ .01)\). However, workshop females showed higher perception score compared to workshop males \((X_{\text{female}} = 4.125; X_{\text{male}} = 4.012)\), see Table 6. It is interesting, however, that the workshop female learners showed significantly higher score than workshop males on monitoring-learning resources characteristic \((t = 2.01, p < .01)\). This could be attributed to higher levels of patience and inquiry abilities in females as compared to males while monitoring learning resources (McGill et al., 2014), see Table 7.

We found these results in line with econometric and sociological models of education proposed by Stage and Hossler (1989) where factors such as parental encouragement, educational aspirations, family income, parental education levels also have shown influence in choice of education, which is very predominantly noticed in female learners, although no direct causality has been established in these studies. We found similar results in our study, as most of our sample come from rural and semi-rural backgrounds.

Table 7. Gender perception analysis- II

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Group</th>
<th>LO</th>
<th>LR</th>
<th>LC</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning</td>
<td>Workshop_males</td>
<td>3.66</td>
<td>3.88</td>
<td>4.42</td>
<td>2.76(LO)</td>
<td>.006(LO)**</td>
</tr>
<tr>
<td></td>
<td>Class_males</td>
<td>3.82</td>
<td>3.875</td>
<td>4.064</td>
<td>1.18(LR)</td>
<td>.237(LR)</td>
</tr>
<tr>
<td></td>
<td>Workshop_females</td>
<td>3.91</td>
<td>3.99</td>
<td>4.03</td>
<td>0.92(LO)</td>
<td>.019(LO)**</td>
</tr>
<tr>
<td></td>
<td>Class_females</td>
<td>4.18</td>
<td>4.61</td>
<td>4.55</td>
<td>2.50(LR)</td>
<td>.018(LO)**</td>
</tr>
<tr>
<td></td>
<td>Overall_males</td>
<td>3.95</td>
<td></td>
<td></td>
<td>4.79</td>
<td>.0022**</td>
</tr>
<tr>
<td>Monitoring</td>
<td>Workshop_males</td>
<td>4.03</td>
<td>4.06</td>
<td>3.93</td>
<td>0.473(LO)</td>
<td>.6359(LO)</td>
</tr>
<tr>
<td></td>
<td>Class_males</td>
<td>4.00</td>
<td>4.07</td>
<td>4.22</td>
<td>0.072(LR)</td>
<td>.9420(LR)</td>
</tr>
</tbody>
</table>
### Table 8. Gender perception analysis—III

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Group</th>
<th>Mean</th>
<th>LO</th>
<th>LR</th>
<th>LC</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Workshop_male</td>
<td>4.03</td>
<td>4.06</td>
<td>3.93</td>
<td>3.80(LC)</td>
<td>.00017(LC)**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Class_female</td>
<td>3.971</td>
<td>3.875</td>
<td>3.75</td>
<td>1.23(LR)</td>
<td>.218(LR)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Workshops_female</td>
<td>3.977</td>
<td>3.875</td>
<td>4.130</td>
<td>0.071(LO)</td>
<td>.942(LO)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Class_male</td>
<td>3.82</td>
<td>4.07</td>
<td>4.22</td>
<td>2.11(LR)</td>
<td>.036(LR)**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Class_female</td>
<td>3.977</td>
<td>3.875</td>
<td>4.130</td>
<td>0.363(LO)</td>
<td>.716(LO)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Workshop_male</td>
<td>4.03</td>
<td>4.06</td>
<td>3.93</td>
<td>0.826(LO)</td>
<td>.4096(LO)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Workshop_female</td>
<td>3.97</td>
<td>3.875</td>
<td>4.130</td>
<td>2.01(LR)</td>
<td>.044(LR)*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Workshops_female</td>
<td>4.00</td>
<td>4.07</td>
<td>4.22</td>
<td>1.32(LR)</td>
<td>.1867(LR)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Class_female</td>
<td>3.971</td>
<td>3.875</td>
<td>3.75</td>
<td>1.131(LC)</td>
<td>.2595(LC)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Workshops_male</td>
<td>4.13</td>
<td>3.88</td>
<td>3.95</td>
<td>2.83(LO)</td>
<td>.0042(LO)**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Class_male</td>
<td>3.93</td>
<td>3.86</td>
<td>4.04</td>
<td>0.2991(LR)</td>
<td>.765(LR)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Class_female</td>
<td>4.13</td>
<td>3.88</td>
<td>3.95</td>
<td>3.52(LO)</td>
<td>.0051(LO)**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Workshops_female</td>
<td>4.05</td>
<td>3.80</td>
<td>3.98</td>
<td>1.977(LO)</td>
<td>.043(LO)*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Class_male</td>
<td>3.93</td>
<td>3.86</td>
<td>4.04</td>
<td>1.29(LO)</td>
<td>.196(LO)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Class_female</td>
<td>4.05</td>
<td>3.80</td>
<td>3.98</td>
<td>0.533(LR)</td>
<td>.594(LR)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Workshops_male</td>
<td>4.13</td>
<td>3.88</td>
<td>3.95</td>
<td>1.19(LO)</td>
<td>.235(LO)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Class_male</td>
<td>3.93</td>
<td>3.86</td>
<td>4.04</td>
<td>0.3411(LR)</td>
<td>.733(LR)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Class_female</td>
<td>4.05</td>
<td>3.80</td>
<td>3.98</td>
<td>0.296(LO)</td>
<td>.767(LC)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Class_male</td>
<td>3.93</td>
<td>3.86</td>
<td>4.04</td>
<td>0.797(LO)</td>
<td>.4262(LO)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Class_female</td>
<td>3.86</td>
<td>3.76</td>
<td>3.98</td>
<td>0.8827(LR)</td>
<td>.3785(LR)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overall_males</td>
<td>4.04</td>
<td>2.22</td>
<td>.026*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overall_females</td>
<td>3.95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.** *p < .05; **p < .01.*
Workshop females vs. classroom males

- Workshop females scored 67.45% while classroom males scored 64.7%. However, overall workshop group scores were higher. Interestingly, workshop females scored higher than classroom males. We also found statistically significant differences among workshop females and classroom males ($t = 2.05, p < .05$), see Table 3. Workshop females seem to benefit more from Spoken Tutorial methodology than classroom males who showed faith in instructor driven approach as compared to self-learning (Allen et al., 2002).

- We also found statistically significant differences among workshop females and classroom males for planning-learning objective, planning-learning resource and planning-learning outcomes ($t = 3.52, p < .01$; $t = 8.75, p < .01$; $t = 5.73, p < .01$). And also for monitoring-learning resources, monitoring-learning outcomes respectively ($t = 2.11, p < .05$; $t = 5.82, p < .01$), see Table 7. However, no such significance was found for adaptability and evaluation characteristics among workshop females and classroom male learners, see Table 8. These results are similar to Murray et al. (2015) who found that structured orientation of course content and activity driven approach has contributed to improvement in student performance. Our methodology also uses activity driven approach, which is a concern in conventional classrooms owing to course deadlines or teaching staff issues (Lee, 2014).

Workshop males vs classroom females

- Workshop males scored 69.9% while classroom females scored 63.95%. Workshop males scored higher than classroom group. We found statistical significance for Java test among workshop males and classroom females ($t = 6.03, p < .01$). Studies in computing education have shown similar differences among the genders as stated by Hasan (2003), which seem to improve gradually with more females’ participation through self-learning methods.

- Perception scores also showed statistically significant differences among the workshop males and classroom females for planning-learning objective ($t = 2.76, p < .01$), planning-learning outcome ($t = 5.72, p < .01$), adaptability-learning objective ($t = 3.52, p < .01$). This could be due to workshop males paid more attention to choice of the course as compared to classroom females who mostly depend on course instructor’s advice.

Classroom males vs classroom females

- Classroom males scored 64.7% while classroom females scored 63.95%. We did not find any statistically significant differences among the classroom males and females ($t = 0.03, p = .972$) for the Java test. However, classroom males performed better than classroom females, see Table 3.

- Although classroom males showed higher perception scores compared to classroom females ($X_{LC} = 4.0, X_{LR} = 4.07, X_{LC} = 4.22$), it is not at all statistically significant in any of the planning, monitoring, adaptability and evaluation characteristics, see Tables 6-8. Perhaps excessive spoon feeding in a classroom setting has made the entire group somewhat homogeneous.

- Schunk et al. (2012) have also found that the self-learning ability of classroom students was low, which is consistent with the findings of the present research. This also answers the third research question: the above said gender differences and learning experiences of the learners influences their self-learning abilities.

Generalization

We believe that the results of this study will extend for most other programming languages as well, going by the widespread adoption of the self-learning methodology through Spoken Tutorials and the feedback received from many (Moudgalya, 2014). The following is a sample testimonial that we received in this regard:
I find the OpenFOAM lecture video tutorials very useful . . . Tutorials in Salome are also my interests, as well as Blender, Scilab, and C++. I am already advocating your website! You are really helping out people, especially those who are not knowledgeable with open source programs, which are essentially the trend nowadays due to economic reasons. Howell Gonzales, graduate student at Kansas State University.

Further, we believe that our methodology is useful in teaching methods that use synchronous and asynchronous teaching strategies. Students can be asked to self-learn through the asynchronous phase, and their doubts can be resolved during synchronous sessions. MOOCs and flipped classroom methods are possible application areas of this approach. Indeed, online platforms, such as (EdX, see http://www.edx.org), have shown that this approach can result in a superior performance, as compared to conventional classroom based teaching methods (Schmid et al., 2014). Preliminary studies carried out by the authors (Eranki & Moudgalya, 2013) also confirm this generalization. Although the proposed method seems to be effective to a restricted class of topics, the benefits can still be enormous, as the number of people to be trained for employment is close to half a billion in India alone (Financial Express Bureau, 2015).

Conclusions

The current study has focused on the comparison of effectiveness of a Spoken Tutorial based self-learning workshop with that of classroom teaching by using a standardized SRL Questionnaire and a Java test. As this approach is scalable, it will become extremely useful, if it is as effective as the conventional approach of teaching. We validated the effectiveness of this method in this work. As a matter of fact, the workshop trained students did better than the students who underwent conventional method, with statistical significance. A large number of self-learners wanted to use Spoken Tutorials for other topics as well, reconfirming the effectiveness of this approach.

The current study has found that there are significant gender differences among the male and female learners. Workshop males have shown high on adaptability, monitoring and planning characteristics as compared to workshop females, whereas, the latter are high on monitoring learning resources. Workshop females have done better than classroom females. More interestingly the former have done better than classroom males also, a surprising result, as generally, males perform better than females on programming topics.

The results of the current study have significant implications to instructional design and course content, especially through the modern content delivery means. This approach is expected to help everyone, but in particular, the females. This study also shows the suitability of the Spoken Tutorial approach for self-learning of programming concepts. Although we restricted this study to Java training only, we believe that this approach is useful for other programming types of courses also, as seen by the testimonials the Spoken Tutorial methodology received (see http://spoken-tutorial.org/testimonials/). We expect this approach to be useful to several other skill based topics as well.

Acknowledgements

This work was partly funded by the NMEICT-MHRD GoI through the Talk to a Teacher project. We thank spoken tutorial project-staff members and the participants of the study for their time and efforts.

References


Barac, R., & Bialystok, E. (2012). Bilingual effects on cognitive and linguistic development: Role of language, cultural background, and education. *Child development, 83*(2), 413–422.


Moudgalya, K. M. (2014). Pedagogical and organisational issues in the campaign for it literacy through spoken tutorials. In R. Huang, Kinshuk, & N. S. Chen (Eds.), *The new development of technology enhanced learning* (pp. 223–244). Berlin Heidelberg: Springer-Verlag.


Effects of WOE Presentation Types Used in Pre-training on the Cognitive Load and Comprehension of Content in Animation-Based Learning Environments

Jaewon Jung*, Dongsik Kim* and Chungsoo Na
Department of Educational Technology, Hanyang University, Seongdong-gu, Seoul, South Korea // jjungj5@gmail.com // kimdsik@hanyang.ac.kr // skcjdttn@gmail.com

*Corresponding authors

(Submitted August 19, 2015; Revised November 19, 2015; Accepted February 1, 2016)

ABSTRACT
This study investigated the effectiveness of various types of worked-out examples used in pre-training to optimize the cognitive load and enhance learners’ comprehension of the content in an animation-based learning environment. An animation-based learning environment was developed specifically for this study. The participants were divided into four groups; each group was provided with a different type of WOE for pre-training: animation with narration; animation with concise narration; animation with narration and concise textual information; and animation with concise narration and verbal labels. After the pre-training phases, the participants’ cognitive load was measured in three dimensions (intrinsic, extraneous, and germane cognitive loads), and their comprehension of the content was analyzed. Next, during the training phase, the participants were asked to carry out a complete task using animation accompanied by narration and concise textual information. After the training phase, the participants’ cognitive load and comprehension of the content were analyzed again. The findings revealed that animation accompanied by concise narration and verbal labels was the most effective in controlling the cognitive load and enhancing the learning experience.

Keywords
Animation, Cognitive-load theory, Worked-out examples, Animation-based learning, Multimedia design

Introduction

Animation is frequently used to enhance learners’ acquisition of complex knowledge. Types of animation can be based on their level of motion: either static (e.g., illustrations and other depictions) or active (e.g., videos) (Mayer, 2002). In this study, the term “animation” refers to the presentation of material in a dynamic form, such as video content. Some researchers have reported that animation-based learning materials are more effective than static depictions (Hegarty & Kriz, 2008; Park & Gittelman, 1992), but others have disagreed (Hegarty, 2004; Tversky, Morrison, & Betrancourt, 2002). Animation-based learning experiences are sometimes ineffective because learners have difficulty managing and integrating visual learning content that is provided dynamically; because successive images often replace each other within seconds, learners may fail to notice important information (Mayer, 2008).

One way to overcome the difficulties involved in the use of animation-based learning materials is to provide pre-training in the basic content of the animations. Pre-training the learners in the key terms and the core learning content of animation-based materials is an important means of optimizing the cognitive load and making learning experiences more successful. In general, previous researchers have tended to use the cognitive-load theory (CLT) to investigate the effectiveness of pre-training based on worked-out examples (WOEs) and to identify productive types of pre-training (Lowe, 2008; Mayer, Mathais, & Wetzell, 2002). Some researchers have reported that providing WOE could reduce the cognitive load (Clarke, Ayres, & Sweller, 2005; Renkl & Atkinson, 2003), whereas others have reported that it could increase the cognitive load due to the expertise reversal effect (Kalyuga, 2014; Kalyuga, Chandler, Tuovinen, & Sweller, 2001). To optimize multimedia-based learning, it is crucial to gain some experience of the learning method—through pre-training, for example—to reduce the cognitive load created by complex animation-based learning tasks. It is not easy to use animation effectively to deliver learning content and enhance learning. Therefore, animation used to present information must be carefully designed, with attention to the learner’s cognitive load and information processing capacity, to ensure that the animation type optimizes the learning experience (Mayer, 2009).

Most previous studies of information presentation in multimedia learning have revealed the effects of animation-based information presentation (Yang, Andre, Greenbowe, & Tibell, 2003). In particular, the effectiveness of static graphics has been compared with that of animation (Hegarty & Kriz, 2008), and types of productive information presentation using animation have been explored. According to Gkins (2006), one effective
Although empirical findings showed that an integrated format was superior to spatial separation of graphics, there is a slightly different perspective in which split-source formats increase the extraneous load and simultaneously increase the germane load (Cierniak, Scheiter, & Gerjets, 2009; Kester, Kirschner, & van Merriënboer, 2005). Although the types of effective information presentation could differ from that in multimedia learning to conduct pre-training, research was not sufficient to explore efficiently the effects of information presentation for pre-training. In addition, systematic exploration of the effects of using animated WOEis to pre-train learners is lacking. Cognitive load theorists have recommended the use of a multimedia design to reduce the extraneous cognitive load (Mousavi, Low, & Sweller, 1995). However, few efforts have explored the use of different types of animation-based information presentation to reduce the extraneous load and simultaneously increase the germane cognitive load (Atkinson, Mayer, & Merrill, 2005; Moreno & Mayer, 2004). Thus, this study was performed to provide empirical evidence on productive types of animation-based information presentation for pre-training that are effective in increasing the germane cognitive load and reducing the extraneous cognitive load beyond optimizing the learning experience.

This study was designed to investigate the effects of WOE presentation types of pre-training in controlling the cognitive load and enhancing the learning performance. We used the following four types of WOE in the pre-training phases, based on the principle that the multimedia design affects the cognitive load: animation with narration (AN); animation with concise narration (ACN); animation with narration and concise text (ANCT); and animation with concise narration and verbal labels (ACNW).

The specific research questions addressed were “What are the effects of WOE types used in pre-training on the learners’ cognitive load in an animation-based learning environment?” and “What are the effects of WOE types used in pre-training on the comprehension of content in the animation-based learning environment?”

Theoretical background

Instructional design to optimize cognitive load

According to the CLT, instruction should be designed to reduce the extraneous cognitive load and increase the germane cognitive load, because learners’ working-memory space is limited (Sweller, van Merriënboer, & Paas, 1998). The CLT studies have proposed various instructional-design strategies to optimize the cognitive load and thereby enhance learners’ automation and schema construction, given the limited capacity of working memory (Paas & Sweller, 2012; Pollock, Chandler, & Sweller, 2002; Renkl & Atkinson, 2003; Yang et al., 2003). An instructional strategy that promotes successful learning by optimizing the learners’ cognitive-load resources—intrinsic, extraneous, and germane—is described as follows.

The intrinsic load is determined by the complexity of the learning task and is created by the process of organizing information in the working memory. In particular, an intrinsic load is generated when the number and interactivity of learning-task components are high and the learner has little relevant prior knowledge. Excessive intrinsic load reduces learners’ motivation and leads to learning failure. Therefore, instructional designers must take learning content complexity and prior knowledge into consideration. The intrinsic load can be regulated by presenting simple and complex learning content (Gerjets, Scheiter, & Catrambone, 2006; Salden, Pass, & van Merriënboer, 2006) accompanied by pre-training to enhance the domain knowledge required to gain task expertise (Clarke et al., 2005; Si, Kim, & Na, 2014).

The extraneous cognitive load is determined by the learning method, the method of information presentation, the time taken to present the learning content, the nature of the learning content, and the structural quality of the learning strategy. The extraneous cognitive load interferes with learning. Instructional principles typically used to reduce the extraneous cognitive load and thereby mitigate its negative effect on learning include the redundancy principle and the contiguity principle (Sweller, 2010).

The germane cognitive load is determined by the cognitive effort required for schema construction and automation; this type of cognitive load is generative and is directly related to learning. The germane cognitive load is created by learning methods designed to promote automation and schema formation in the working memory by reducing the learner’s extraneous cognitive load with the use of cognitive activities such as elaboration, abbreviation, comparison, and inference (Moreno & Park, 2010). Researchers have generally argued that the germane cognitive load secures space for schema acquisition by means of instructional designs that reduce extraneous cognitive load. However, it has also been suggested that the space secured for the germane
cognitive load is proactively used by the learner for schema acquisition, eliciting increasing interest in the development of instructional-design strategies that enhance the germane cognitive load.

**Multimedia learning**

The multimedia-learning strategy proposed by Mayer (2008) minimizes extraneous processing by limiting working-memory capacity to essential processing, thus securing space for automated schema construction. In addition, multimedia learning accelerates generative processing and thereby enhances passive processing without exceeding the learner’s available cognitive capacity. Mayer (2008) demonstrated that learners process multimedia learning content within their limited cognitive capacity. In the following sections, extensions of the multimedia-design principle to the information presentation types addressed in this study are discussed.

**Modality principle**

The modality principle concerns the enhancement to learning that results from using graphics and narration rather than graphics and text to present learning content, because the former enable the content to be processed separately through verbal and visual channels (Low & Sweller, 2005). In this study, the modality principle was applied to all four information presentation types (AN, ACN, ANCT, and ACNW) to solve problems that arise from the complexity of the learning content and to moderate the learners’ intrinsic cognitive load. In general, the findings of related studies have shown that the simultaneous use of graphics and narration to present information has a more positive influence on learning than the simultaneous use of graphics and text. WOEs that are based on AN or ACN allow the learning content to be processed separately through verbal and visual channels, thereby reducing the cognitive load. Therefore, these types of WOE are likely to be particularly useful when the learning content is complex and rapidly presented.

**Multimedia principle**

The multimedia principle concerns the enhancement generated when text and graphics are presented together, compared with the use of text alone. This principle was applied to all four of the information presentation types used in this study. The multimedia-design principle has been used to increase the germane cognitive load. “Text” is assumed to refer to printed text or narration composed of words, whereas the term “graphics” describes static forms such as illustrations and other images, along with dynamic forms such as video clips. Learning content to which the multimedia principle has been applied is presented in the form of both text and graphics to engage and integrate the learners’ verbal and visual mental modes. This type of learning content is most suitable for use with learners with relatively little relevant knowledge.

**Redundancy principle**

Because the AN and ACN WOEs minimize the extraneous cognitive load, they are consistent with the redundancy principle, a component of the multimedia-design principle. The redundancy principle describes the positive effects on learning of presenting content with graphics and narration only, rather than with graphics, narration, and text. This distinction is due to learners’ limited cognitive capacity; the textual learning content increases the cognitive load in the visual-processing channel and decreases the learners’ capacity to process animation-based content. Chandler and Sweller (1991) found that when content that can be sufficiently explained with graphics is presented in both visual and textual form, redundant information must be removed to enhance learning, because learners are already using their visual channel to process the graphics. Therefore, both AN WOE and ACN WOEs were expected to reduce the extraneous cognitive load because they do not require the learner to process the animation and text presented on the screen simultaneously.

**Spatial-contiguity principle**

The spatial-contiguity principle is a multimedia-design principle applied in this study to the ACNW WOE. This principle minimizes cognitive processing by positioning related graphics and text in close proximity to ensure that the learner’s attention is not divided during learning. The spatial-contiguity principle can be used to reduce the extraneous cognitive load generated by an inappropriate layout. In a study conducted in a multimedia-
learning environment, Moreno and Mayer (1999) found that positioning text near related graphics had more positive learning outcomes than spatial separation of graphics from the related text at the bottom of the screen. The results of this study were consistent with the spatial-contiguity principle, according to which the presentation of related information in close proximity can reduce the extraneous cognitive load.

**Methods**

**Participants**

Ninety-two students at a 4-year college who had enrolled in one or more of four educational technology classes were invited to participate in the study. The participants consisted of 58 women and 34 men, all of whom were juniors or seniors. Their ages ranged from 20 to 27 years (mean, 23.12 years), and their majors ranged from Korean to mathematics education, educational technology, and practical art. When the students logged on to the website developed for the experiment, they were randomly assigned by the system to one of four groups (resulting in 23 students per group). The students in the first group were given an AN WOE; the second group received an ACN WOE, the third group received an ANCT WOE; and the fourth group received an ACNW WOE. The data obtained from all 92 students were prepared for statistical analysis.

**Experimental materials**

*Use of four types of WOE in pre-training*

During their pre-training, the participants learned how to replace a car tire. The desired exit behavior was a demonstration of their understanding of and ability to accurately explain the steps involved in replacing a car tire. The four types of WOE were constructed as shown in Figure 1: AN, ACN, ANCT, and ACNW.

![Figure 1. Example frames from animation-based WOEs for pre-training](image)

**Pretests**

Three computer-based tests (terminology, labeling, and simple facts) were used to measure the participants’ prior knowledge. Tests consisted of five “true or false” problems and five multiple-choice problems (see Table 1). Two computer-based tests (paper folding and mental rotation) were used to measure the participants’ spatial ability. Each test consisted of 10 multiple-choice problems (see Figure 2).

![Figure 2. Sample pre-test items to measure spatial ability](image)
Table 1. Sample pre-test items to measure prior knowledge

<table>
<thead>
<tr>
<th>Item #</th>
<th>Problem</th>
<th>True</th>
<th>False</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A spare tire is stored on the floor of the trunk.</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>2</td>
<td>A wrench is required to unfasten a tire bolt.</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>3</td>
<td>A jockey is a tool that can lift the car body.</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

Cognitive-load measures

A seven-point Likert scale was developed to measure mental effort, based on previous research by Cierniak et al. (2009), Jackson and Marsh (1996), and Ryu and Yim (2009) (see Table 2). The participants were asked to rate their perceived mental effort on a scale ranging from “extremely easy” (1) to “extremely difficult” (7) (see Figure 3). The ratings for both initial and final subjective mental effort had exactly the same underlying structure, and each subjective mental effort rating was given in response to the same nine questions. The rating was given two times: after the pre-training phase and the training phase. The reliability analysis revealed a Cronbach’s alpha value of 0.73.

Table 2. Sample questions for mental-effort rating

<table>
<thead>
<tr>
<th>Type of load</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intrinsic load</td>
<td>The learning content was difficult (Cierniak, Scheiter, &amp; Gerjets, 2009).</td>
</tr>
<tr>
<td>Extraneous load</td>
<td>It was difficult to use the material to learn (Ryu &amp; Yim, 2009).</td>
</tr>
<tr>
<td>Germane load</td>
<td>I concentrated while learning (Jackson &amp; Marsh, 1996).</td>
</tr>
</tbody>
</table>

Figure 3. Mental effort rating scale

Diagnostic tests

Diagnostic tests were used to measure the participants’ comprehension of the learning content. Each test consisted of 10 narrative-recall problems. The participants took a diagnostic test after the pre-training phase and after the training phase (see Table 3).

Table 3. Sample diagnostic-test items

<table>
<thead>
<tr>
<th>Step</th>
<th>Content</th>
<th>Detailed description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Opening the trunk and checking for a spare tire and equipment.</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>Separating spare tire.</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>Removing tools.</td>
<td>-</td>
</tr>
</tbody>
</table>

Animation-based learning environment

The structure of the animation-based learning environment developed for use in this study is presented in Figure 5. The students carried out seven sets of learning tasks: pre-tests, pre-training, an initial mental-effort rating, initial diagnostic tests, training, a final mental-effort rating, and final diagnostic tests (see Figure 4).

Figure 4. Learning processes in animation-based learning environment
Procedure

The experiments were conducted in the form of an assignment that lasted 1 week. The participants were first provided with the uniform resource locator of the animation-based learning environment in class. The students participated individually in the experiment. None of the phases were time limited, but the participants were instructed to undertake the program only once per 2 hours. An instructor provided the participants with special instructions for completing the assignment in class. After the pre-test, the participants were asked to study the terminology of cars and car tires using the same WOE type: a static form that combined text with pictures. Next, the groups received different types of WOE to provide the procedural information on replacing a tire. The participants were allowed to control the exercise by pressing the “Pause” or “Play” buttons during the pre-training phase and the training phase. After completing the pre-training, each participant provided an initial mental-effort rating and took a diagnostic test. Soon after the diagnostic test, the participants were asked to carry out a complete task using the same WOE, combining animation, narration, and concise text, for training. After completing the training phase, all of the participants provided final mental-effort ratings and took diagnostic tests.

Data collection and analysis

The independent variables under study were the four different types of WOE, and the dependent variables were mental effort and comprehension of the content. The learners’ prior knowledge and spatial ability were used as covariates to test the flawless effect of pre-training. The participants’ comprehension of the content was computed from their scores on the diagnostic test. An analysis of covariance conducted with PSAW Statistics 18 was used to compare the mental effort and comprehension of the content exhibited by the four groups. The significance level for statistical verification was set at 0.05.

Results and discussion

Effects of WOE types used in pre-training on learners’ cognitive load

During pre-training

The ANCT groups had the greatest intrinsic load during the pre-training phase, and the ACN group had the lowest intrinsic load. The ACNW group had the lowest extraneous load and the greatest germane load (see Table 4). The results of the analysis of covariance for prior knowledge and spatial ability revealed that prior knowledge had a statistically significant effect on the intrinsic load \(F(1, 86) = 8.59, p = .004\) (see Table 5). However, no significant differences were seen between the extraneous load and the germane load \(F(1, 86) = 0.05, p = .825, F(1, 86) = 0.06, p = .808\). When controlling for spatial ability, no significant differences were found among the intrinsic load, the extraneous load, and the germane load \(F(1, 86) = 0.02, p = .881, F(1, 86) = 0.08, p = .775, F(1, 86) = 0.13, p = .716\). However, the adjusted means for the intrinsic load, the extraneous load, and the germane load differed significantly among the four groups when controlling for prior knowledge and spatial
The WOE based on ACNW was found to most effectively reduce the extraneous load and increase the germane load, consistent with the multimedia principle, the modality principle, and the spatial-contiguity principle. Chandler and Sweller (1991) and Sweller and Chandler (1994) found that presenting verbal and visual explanations together is more effective than presenting a verbal or visual explanation alone, but that the redundant presentation of verbal and visual explanations is less effective than the use of a single medium. When information that demands visual processing is presented redundantly along with text, cognitive overload occurs in the visual channel and the learner experiences difficulty in integrating the information. Likewise, in this study, ACNW had the possibility to increase the extraneous load based on the redundancy principle because it was presented as animation with concise narration along with verbal labels. However, although Kalyuga and Sweller (2014) reported that redundant graphics are not effective in learning, it was found that verbal labels had no influence on the extraneous load in this study. Overall, ACNW was found to be the most effective in optimizing mental effort for pre-training.
During training

In terms of the effects on training of the WOE types used in pre-training, the AN group had the greatest intrinsic and extraneous loads, and the ACNW group had the lowest intrinsic and extraneous loads. The germane load generated in the ANCT group was the highest (see Table 6). When prior knowledge and spatial ability were treated as covariates, prior knowledge was found to have a statistically significant effect on the intrinsic cognitive load \( F(1, 86) = 4.40, p = .039 \) (see Table 7). However, no significant differences were found between the extraneous load and the germane load \( F(1, 86) = 1.09, p = .299, F(1, 86) = 0.75, p = .390 \). No significant differences were found among the intrinsic load, the extraneous load, and the germane load for the spatial-ability covariate \( F(1, 86) = 0.23, p = .630, F(1, 86) = 0.06, p = .809, F(1, 86) = 1.64, p = .204 \). In contrast, the adjusted means obtained for the extraneous load differed significantly among the four groups when controlled for prior knowledge and spatial ability \( F(3, 86) = 3.11, p = .031 \). However, no significant differences were seen in the adjusted means of either the intrinsic load or the germane load \( F(1, 86) = 1.35, p = .265, F(1, 86) = 0.75, p = .527 \).

<table>
<thead>
<tr>
<th>WOE type</th>
<th>Intrinsic load</th>
<th>Extrinsic load</th>
<th>Germane load</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( M )</td>
<td>( SD )</td>
<td>( AM )</td>
</tr>
<tr>
<td>AN</td>
<td>5.91</td>
<td>3.20</td>
<td>5.87</td>
</tr>
<tr>
<td>ACN</td>
<td>5.04</td>
<td>3.25</td>
<td>5.14</td>
</tr>
<tr>
<td>ANCT</td>
<td>5.48</td>
<td>2.61</td>
<td>5.51</td>
</tr>
<tr>
<td>ACNW</td>
<td>4.35</td>
<td>2.33</td>
<td>4.27</td>
</tr>
</tbody>
</table>

Note. \( N = 23 \). \( M \) = Mean. \( SD \) = Standard deviation. \( AM \) = Adjusted mean. AN = Animation with narration. ACN = Animation with concise narration. ANCT = Animation with narration and concise text. ACNW = Animation with concise narration and verbal labels.

<table>
<thead>
<tr>
<th>Type of load</th>
<th>Source</th>
<th>( SS )</th>
<th>( df )</th>
<th>( MS )</th>
<th>( F )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intrinsic load</td>
<td>Model</td>
<td>73.97</td>
<td>5</td>
<td>14.79</td>
<td>1.86</td>
<td>.110</td>
</tr>
<tr>
<td></td>
<td>Prior knowledge</td>
<td>35.06</td>
<td>1</td>
<td>35.06</td>
<td>4.40</td>
<td>.039</td>
</tr>
<tr>
<td></td>
<td>Spatial ability</td>
<td>1.86</td>
<td>1</td>
<td>1.86</td>
<td>0.23</td>
<td>.630</td>
</tr>
<tr>
<td></td>
<td>WOE type</td>
<td>32.15</td>
<td>3</td>
<td>10.72</td>
<td>1.35</td>
<td>.265</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>684.51</td>
<td>86</td>
<td>7.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>758.48</td>
<td>91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extrinsic load</td>
<td>Model</td>
<td>55.87</td>
<td>5</td>
<td>11.17</td>
<td>2.11</td>
<td>.072</td>
</tr>
<tr>
<td></td>
<td>Prior knowledge</td>
<td>5.79</td>
<td>1</td>
<td>5.79</td>
<td>1.09</td>
<td>.299</td>
</tr>
<tr>
<td></td>
<td>Spatial ability</td>
<td>0.31</td>
<td>1</td>
<td>0.31</td>
<td>0.06</td>
<td>.809</td>
</tr>
<tr>
<td></td>
<td>WOE type</td>
<td>49.43</td>
<td>3</td>
<td>16.48</td>
<td>3.11</td>
<td>.031</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>455.82</td>
<td>86</td>
<td>5.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>511.69</td>
<td>91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Germane load</td>
<td>Model</td>
<td>31.54</td>
<td>5</td>
<td>6.31</td>
<td>0.81</td>
<td>.549</td>
</tr>
<tr>
<td></td>
<td>Prior knowledge</td>
<td>5.84</td>
<td>1</td>
<td>5.84</td>
<td>0.75</td>
<td>.390</td>
</tr>
<tr>
<td></td>
<td>Spatial ability</td>
<td>12.82</td>
<td>1</td>
<td>12.82</td>
<td>1.64</td>
<td>.204</td>
</tr>
<tr>
<td></td>
<td>WOE type</td>
<td>17.54</td>
<td>3</td>
<td>5.85</td>
<td>0.75</td>
<td>.527</td>
</tr>
<tr>
<td></td>
<td>Error</td>
<td>673.11</td>
<td>86</td>
<td>7.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>704.65</td>
<td>91</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. \( N = 23 \). \( SS \) = Sum of squares. \( MS \) = Mean square.

Although the results were not statistically supported, the ACNW group exhibited lower intrinsic and extraneous loads and a greater germane load during the pre-training and the training, as shown in Table 6. In addition, based on the results, the AN group had the highest extraneous load despite its consistency with the redundancy principle, whereas the ACNW group had the lowest extraneous load. The findings provide an extension of the spatial-contiguity principle, in which the keywords are positioned next to the graphics to avoid splitting the attention of the learner, thereby minimizing the cognitive load that can occur when integrating visual and verbal information. Although the verbal labels in the narration were redundant, this information presentation type carried little risk of increasing the extraneous cognitive load, due to the spatial-contiguity principle. In addition, because the AN group had the greatest extraneous load for the pre-training phase and the training phase, it was expected that the WOE based on AN would not be the most suitable type of WOE for learning content that provides procedural information. Based on the results, we can propose that the characteristics of the learning content should be considered to lead to a successful learning experience.
Of the four WOE types used in the pre-training phase, ACNW was found to be the most effective in reducing the intrinsic and extraneous loads during the training phase. However, ANCT was found to be the most effective in accelerating the development of the germane cognitive load during training. This finding implies that learners who received the same types of information presentation during the pre-training phase and the training phase experienced more efficient generative cognitive processing. Although ANCT was found to be the most effective method in increasing the germane load, ACNW was also found to be effective in generating the germane load and the mean differences among the groups were not high. Based on the result, it was expected that the ACNW group experienced a lower extraneous load and secured working-memory space for the germane load, resulting in an increase in the germane load (Sweller, Ayres, & Kalyuga, 2011).

The intrinsic load and extraneous load generated during the training phase were lower than those generated during pre-training, but the germane load was greater. This result is consistent with the observed effects of the WOE types used in pre-training on the learners’ cognitive load and the effects of prior knowledge on the learners’ domain expertise and cognitive load. In addition, the finding is line with those of previous studies that demonstrated the effects of pre-training within CLT (Lowe, 2008; Mayer et al., 2002). Increasing learners’ relevant expertise with pre-training helps them to build cognitive schema and thus reduces their intrinsic load. Kalyuga (2014) reported that an instructional method that is effective for learners with less prior knowledge is ineffective for learners with more prior knowledge. However, the result suggests that pre-training positively influenced training and that there were no effects of expertise reversal. Therefore, pre-trained learners are likely to be better able to focus on the key information provided in verbal and visual forms during learning. This inference supports the findings of previous CLT-based studies on the complexity of learning tasks and the level of learner expertise (Mayer & Moreno, 2003; Renkl & Atkinson, 2003).

Effects of WOE types used in pre-training on comprehension of content

During pre-training

The ACNW group exhibited the greatest comprehension of the learning content during the pre-training phase, whereas the ANCT group showed the least comprehension (see Table 8). When prior knowledge and spatial ability were treated as covariates, no statistically significant differences were found among the groups [$F(1, 86) = 1.43, p = .235, F(1, 86) = 0.02, p = .892$]. However, the adjusted means obtained for content comprehension differed significantly among the four groups when controlling for prior knowledge and spatial ability [$F(3, 86) = 20.72, p < .001$]. The results revealed that the briefer WOEs—ACN and ACNW—had a greater positive influence on the learners’ comprehension of the content.

During training

As in the pre-training test, the ACNW group received the highest score for comprehension of the content during the training phase. The mean differences between the groups were small but were statistically supported (see Table 8). The covariates of prior knowledge and spatial ability did not differ significantly [$F(1, 86) = 0.81, p = .370, F(1, 86) = 0.00, p = .981$] (see Table 9). However, the adjusted means for the comprehension of the content differed significantly among the four groups when controlling for prior knowledge and spatial ability [$F(3, 86) = 9.37, p < .001$].

The use of WOEs affected the learners’ comprehension of content during the pre-training phase and the training phase. In particular, as expected, higher levels of schema acquisition were observed during the training phase than during pre-training, probably because the pre-training improved learner expertise, which in turn increased schema acquisition during the training phase. In addition, the mean differences among the groups during the training phase were smaller than those during pre-training. This implies that the learners gained expertise in their respective tasks during pre-training, which led to better performance during the training phase. This result supports the previous study on the advance organizer (Ausubel, 1960), which reported that meaningful learning can occur via an advance organizer that is relevant with the learners’ prior knowledge. In addition, the advance organizer ensures that meaningful learning is performed by appropriately connecting the newly presented content and the learners’ cognitive structure. Therefore, in this study, it is interpreted as that pre-training contributed to stimulate the prior knowledge of the learners via the successful implementation of the role of the advance organizer.
Table 8. Effects of WOE types on comprehension of content

<table>
<thead>
<tr>
<th>WOE type</th>
<th>Pre-training</th>
<th>Training</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>AN</td>
<td>8.70</td>
<td>0.88</td>
</tr>
<tr>
<td>ACN</td>
<td>8.83</td>
<td>0.72</td>
</tr>
<tr>
<td>ANCT</td>
<td>7.70</td>
<td>0.47</td>
</tr>
<tr>
<td>ACNW</td>
<td>9.43</td>
<td>0.90</td>
</tr>
</tbody>
</table>

Note. N = 23. M = Mean. SD = Standard deviation. AM = Adjusted mean. AN = Animation with narration. ACN = Animation with concise narration. ANCT = Animation with narration and concise text. ACNW = Animation with concise narration and verbal labels.

Table 9. Covariance analysis of comprehension of content by WOE types

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>T</th>
<th>df</th>
<th>PT</th>
<th>MS</th>
<th>T</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>36.70</td>
<td>9.81</td>
<td>5</td>
<td>5</td>
<td>7.34</td>
<td>1.96</td>
<td>12.66</td>
<td>.000</td>
</tr>
<tr>
<td>Prior knowledge</td>
<td>0.83</td>
<td>0.27</td>
<td>1</td>
<td>1</td>
<td>0.83</td>
<td>0.27</td>
<td>1.43</td>
<td>0.81</td>
</tr>
<tr>
<td>Spatial ability</td>
<td>0.01</td>
<td>0.00</td>
<td>1</td>
<td>1</td>
<td>0.01</td>
<td>0.00</td>
<td>0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>WOE type</td>
<td>36.03</td>
<td>9.42</td>
<td>3</td>
<td>3</td>
<td>12.01</td>
<td>3.14</td>
<td>20.72</td>
<td>9.37</td>
</tr>
<tr>
<td>Error</td>
<td>49.85</td>
<td>28.85</td>
<td>86</td>
<td>86</td>
<td>0.99</td>
<td>0.34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>86.55</td>
<td>38.65</td>
<td>91</td>
<td>91</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. N = 23. SS = Sum of squares. MS = Mean square. PT = During pre-training. T = During training.

Of the four types of WOE used in pre-training, the WOE based on ACNW had the greatest positive influence on the comprehension of the content during the pre-training phase and the training phase. The findings of this study provide evidence that ACNW could be effective in preventing the duplication of redundant information and in highlighting key points using verbal labels adjacent to the relevant graphics. This type of WOE was also found to be most effective in reducing the extraneous load; therefore, the findings provide evidence that the information presentation type with the lower extraneous load is also effective in improving the comprehension of the content.

Conclusions and limitations

The results were consistent with the empirical evidence for the multimedia principle previously obtained in CLT research (Mayer, 2009; Mayer, 2014). The WOE based on ACNW used in pre-training was found to be the most effective in optimizing mental effort and enhancing the comprehension of the content, consistent with the multimedia principle, the modality principle, and the spatial-contiguity principle. Pre-training was found to be an effective instructional strategy for optimizing the cognitive load and promoting schema acquisition in an animation-based learning environment. In addition, this study added evidence that pre-training can lead to effective learning because it helps to optimize the cognitive load and promote the learners’ comprehension of the content, particularly as measured by the cognitive load and schema acquisition. These findings yield some implications for further study. First, we identified a multimedia-design strategy that improves learners’ ability to select, organize, and integrate schema by adjusting the intrinsic cognitive load and the extraneous cognitive load and by accelerating the development of the germane cognitive load. Future researchers are advised to extend this study by exploring possible strategies for efficient extraction of information stored in the long-term memory. Second, the WOE based on ACNW had the possibility to increase the extraneous load based on the redundancy principle; however, the information presentation type was not affected by redundancy effect. Therefore, in future studies the issue of the redundancy principle should be expanded to different types of WOEs and to different learning content.

This study may have been limited by its structure. First, the experiment was conducted online in the form of a college assignment over 1 week, and the learners participated individually. Although the instructor provided special instructions in class for carrying out the assignment, the students’ multimedia learning may have been interrupted. Nevertheless, the findings of the study provide insight into the effectiveness of various types of WOE used in pre-training to optimize the cognitive load and enhance learners’ comprehension of the content. Second, learning can be affected by various elements such as synchronization, pauses, and interactions that were not considered in this study. It seems necessary to explore the effects of interactions in further studies to lead to successful learning experiences.
Acknowledgements

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

References


Which Teaching Strategy is Better for Enhancing Anti-Phishing Learning Motivation and Achievement? The Concept Maps on Tablet PCs or Worksheets?

Jerry Chih-Yuan Sun1* and Kuan-Hsien Lee2

1Institute of Education, National Chiao Tung University, Hsinchu, Taiwan // 2Taipei Municipal Wuchang Junior High School, Taipei, Taiwan // csun@nctu.edu.tw // can@mail2000.com.tw

*Corresponding author

(Submitted April 13, 2015; Revised November 24, 2015; Accepted January 26, 2016)

ABSTRACT

The purpose of this study is to evaluate the feasibility of the integration of concept maps and tablet PCs in anti-phishing education for enhancing students’ learning motivation and achievement. The subjects were 155 students from grades 8 and 9. They were divided into an experimental group (77 students) and a control group (78 students). To begin with, the two groups received identical anti-phishing training: the teacher explained the concept of anti-phishing and asked the students questions; the students then used tablet PCs for polling and answering the teachers’ questions. Afterwards, the two groups performed different group activities: the experimental group was divided into smaller groups, which used tablet PCs to draw concept maps; the control group was also divided into groups which completed worksheets. The study found that the use of concept maps on tablet PCs during the anti-phishing education significantly enhanced the students’ learning motivation when their initial motivation was already high. For learners with low initial motivation or prior knowledge, the use of worksheets could increase their posttest achievement and motivation. This study therefore proposes that motivation and achievement in teaching the anti-phishing concept can be effectively enhanced if the curriculum is designed based on the students’ learning preferences or prior knowledge, in conjunction with the integration of mature and accessible technological media into the learning activities. The findings can also serve as a reference for anti-phishing educators and researchers.

Keywords

Anti-phishing education, Tablet PCs, Concept map, Polling, Motivation

Introduction

The purpose of this study is to evaluate the feasibility of the integration of concept maps and tablet PCs into anti-phishing education for enhancing students’ learning motivation and achievement. Phishing is a kind of attack whereby criminals trick Internet users into providing their personal information. According to statistics provided by the U.S. Federal Bureau of Investigation (FBI), losses arising from identity theft, prepaid card fraud, etc., amounted to US$485 million in the United States in 2011. Previous studies have claimed that users are not interested in learning what and how to anti-phish until they get phished (attacked) (Cranor, 2008; Kumaraguru et al., 2007b). Therefore, the purpose of this study was to develop sound strategies for enhancing learners’ motivation to learn about anti-phishing so that they can learn before they get phished, and ultimately the number of victims can be reduced.

Existing anti-phishing education is mostly carried out through contextual training (e.g., written narratives), generally online (Kumaraguru et al., 2007b); however, most people are unwilling to spend time reading anti-phishing documents, so the display of this information on websites or learning materials often yields very limited results (Sheng et al., 2007). Further, this existing anti-phishing education mostly consists of passive learning, in that students browse the websites and learn independently without a dedicated curriculum or articulated teaching-learning approach; therefore, to achieve better results, this study seeks to propose a set of strategies for anti-phishing education that can be applied in a classroom setting. Furthermore, most of the existing anti-phishing studies are focused on teaching university students (Jansson & von Solms, 2013; Yang, Tseng, Lee, Weng, & Chen, 2012). To target lower age group learners, role-play was used to impart education on Internet information safety and investigate whether the learners remembered key safety measures (Wishart, Oades, & Morris, 2007). It was found that further research is needed to analyze courses designed for teenagers as the target group and to impart anti-phishing knowledge in a structured manner.

The 2012 and 2013 Horizon Reports indicated that tablet PCs have seen increasing use as a medium of education recently (Johnson, Adams, & Cummins, 2012; Johnson et al., 2013). Mock (2004) proposed that tablet PCs are an effective tool for assessing students’ performance, preparing lessons, and giving in-class presentations. Compared to personal computers (PCs) and smartphones, the size of tablet PCs better suits small group in-class
learning. This is because tablet PCs are more portable than PCs, and the larger screen size of tablet PCs (compared to smartphones) allows multiple users to share them. They may also provide new opportunities for students and their teachers to interact; however, even if this is the case, an organized communication process for questions and answers during the knowledge construction process will be needed for the effective consolidation of knowledge. Therefore, this study adopts the concept map learning strategy proposed by studies working from the perspective of constructivism (Roth & Roychoudhury, 1993), with an integrated, interactive real-time feedback-based system on tablet PCs for the curriculum design. The hope is that, compared to the traditional curriculum, this approach will more effectively enhance students’ learning motivation and knowledge of phishing and how to handle anti-phishing behaviors, and thus mitigate the damage caused by phishing to individuals and to society as a whole.

In conclusion, with regard to anti-phishing education, a low willingness to learn was commonly observed among students, resulting in the limited effectiveness of existing educational websites and materials (Cranor, 2008; Kumaraguru et al., 2007b; Sheng et al., 2007). In addition, there has been research on anti-phishing education with university students as the target group (Jansson & von Solms, 2013; Yang et al., 2012), and also on educating children on Internet information safety (Wishart et al., 2007). Nonetheless, classroom courses on anti-phishing for teenagers and children that involve peer interaction and concept mapping are still at an exploratory stage of research. Therefore, this research combined tablet technology with concept mapping, and integrated it into an anti-phishing course for teenagers. It is hoped that through interaction and concept mapping, students will be motivated to learn anti-phishing skills, and their knowledge on the topic will deepen. The issues discussed in this study are as follows:

- Whether the experimental group (in-class learning with concept maps on tablet PCs) could more effectively enhance students’ anti-phishing learning motivation than the control group (in-class learning with worksheets).
- Whether the experimental group could more effectively enhance students’ anti-phishing learning achievement than the control group.

**Literature review**

**Phishing and anti-phishing education**

Phishing is a kind of attack whereby criminals trick Internet users into providing their personal information. Phishing makes use of social engineering techniques and technical subterfuge to steal information relating to a consumer’s personal identity and financial records (APWG, 2012). Prevention is always better than cure because even the most robust system has loopholes (Kumaraguru et al., 2009). Therefore, educating Internet users in anti-phishing knowledge is the best way to resolve this issue at its core. Existing anti-phishing education can be classified as (1) contextual training, such as Microsoft’s phishing education page (Microsoft, 2013); (2) game-based training, such as Carnegie Mellon’s Phish Guru (Kumaraguru et al., 2009; Kumaraguru et al., 2007a) or the Anti-Phishing Education Game, where a soldier has to avoid phishing attacks by means of identifying whether the hyperlink is a phishing URL or not (Yang et al., 2012); or (3) embedded training (Kumaraguru et al., 2007b).

Most existing anti-phishing education consists of web-based passive learning in which students browse and learn on their own, without any structured curriculum or learning goals. If learners have the opportunity to get hands-on learning of and exposure to phishing-related knowledge in a classroom setting, their learning should improve (Kumaraguru, Sheng, Acquisti, Cranor, & Hong, 2010). Given this, the emphasis of this study is on designing a curriculum that can stimulate students’ learning motivation for effective anti-phishing education.

**Learning motivation**

Learning motivation is an important factor influencing (maintaining and directing) learning behavior and learning outcomes (Schunk, Meece, & Pintrich, 2013). Learning motivation is an endogenous factor of learning behavior, and its factor construct includes three components. The first is the value component, which reflects an individual’s cognition of the value of and reason for completing the task. The second is an expectancy component, which refers to an individual’s feeling of efficacy and expectations about completing the task. The third is the affective component, which can be understood as the emotional reaction an individual has to a task, which is in fact most likely to be anxiety (Pintrich & Schunk, 1996; 2002).
Learning relies on both intrinsic motivation (e.g., internal needs) and extrinsic motivation (e.g., material rewards) (Benabou & Tirole, 2003). Teachers can directly improve students’ learning motivation through manipulation of controllable factors such as teaching methods, curriculum, and teaching strategies (Urdan & Schoenfelder, 2006). Therefore, this study seeks to enhance students’ motivation to learn about anti-phishing by integrating tablet PCs as an interactive feedback technology, and concept maps into in-class education.

Application of the worksheets and concept map in anti-phishing education

It can be seen from the above discussion that the greatest barrier to anti-phishing education is students’ low willingness to learn (Arachchilage & Asanka, 2012; Tseng, Chen, Lee, & Weng, 2011). Therefore, this section focuses on teaching methods for improving interactions between teachers and students, as well as strengthening the collaborative relationship to facilitate improvements in learning motivation: namely, the tablet PC, used as an interactive feedback technology, and the concept map. If tablet PCs are allocated to individuals or groups of students and used for idea-sharing across an entire class, in-class collaborative learning can be improved, and an atmosphere of active participation fostered (Anderson et al., 2004). Past literature has mentioned the importance of timely feedback such as that which can be provided through tablet PCs, stating that it not only promotes more meaningful interaction between students and materials (Buchanan, 2000), but also effectively improves and maintains the development of student knowledge (Yorke, 2001).

In addition to feedback, the concept map is another method commonly used in education in recent years. Novak, Bob Gowin, and Johansen (1983) developed a set of education methods using the concept map based on Ausubel, Novak, and Hanesian’s (1968) constructivist theoretical framework. Conducting a summary review at the end of a course involves the application of concept maps (Cliburn Jr., 1987). There has been research on combining concept mapping with educational technology as well as on teaching elementary school children (Hwang, Wu, & Kuo, 2013; Hwang, Wu, & Ké, 2011; Yang, Hwang, Hung, & Tseng, 2013). Hwang et al. (2011) showed that combining mobile learning activities with an interactive concept map-oriented approach and conducting exploratory learning outside classrooms could significantly improve children’s learning attitudes and efficacy as well as generate strong interest in learning and active participation. It was seen that combining educational technology with concept mapping and designing interactive teaching activities improved the students’ learning performance and participation in the courses. Worksheets are another frequently used tool in education, and teachers use them to design questions or tasks with various orientations (such as blank-filling, drawing, or open-ended questions) for students to complete. Worksheets could also serve as supplementary materials for knowledge outside the curriculum; by completing tasks on worksheets, learners could proactively construct knowledge. Furthermore, worksheets can also be a tool for assessing learning outcomes (Lee, 2014). Worksheets are often used in educational activities taking place in natural history museums, science museums, and science centers (Science Communicating Venues or SCVs). Teachers see worksheets as a way to keep learners focused on task completion during activities in SCVs. Learners usually complete worksheets orally or in written form, because it is a convenient way for quantitative calculations on task completion (Kisiel, 2003).

To summarize, a teaching format combining concept maps and tablet PCs facilitates co-operation and interaction between teachers and students, and structural tasks on worksheets have a scaffolding effect which helps learners construct knowledge and maintain attention; that is to say, the two formats have different characteristics. Apart from analyzing the low willingness of students to learn anti-phishing skills (Arachchilage & Asanka, 2012; Tseng et al., 2011), most of the existing anti-phishing studies focused on teaching university students (Jansson & von Solms, 2013; Yang et al., 2012). In one study focusing on learners aged 9–12, Wishart et al. (2007) educated the subjects on Internet information safety through online role-play. They found that the students’ motivation for learning could be strengthened, and that there is scope for designing systematic and interactive anti-phishing courses for teenagers and children. Therefore, in this study, we chose concept maps and worksheets to design in-class learning activities and then compare the effects of applying these two tools in anti-phishing education. The objective of the anti-phishing course proposed in this study is that, through a combination of concept mapping and tablet computers, via peer discussion and drawing by hand, the teenagers would deepen their anti-phishing knowledge and actively participate in this course.

The effects of the worksheets and concept maps on learning motivation and achievement

Previous studies have shown that the concept map education method is an effective learning aid (Moen & Boersma, 1997). Through structural hints, learners could extract concepts from what they have learned, which is an effective method of lowering the cognitive load (Sweller, 1994). Liu and Lee (2013) found that the use of the
concept map knowledge management system had a positive and proactive influence on learners of biological science. Completing a concept map in a group could strengthen learning motivation while reducing learning anxiety, as well as lessening the embarrassment of asking questions in class. Liu (2011) found that for learners with below-average writing skills, the learning performance and motivation when creating a concept map via cooperation were superior to that when completing the map alone. In contrast, for learners with better writing skills, the learning outcomes were better when they completed the concept map alone instead of in a group. Ojima (2006) stated that the use of concept maps can benefit learners directly, and is positively correlated to learning experience, motivation, and performance. Kinchin and Hay’s (2005) study showed that for learners with poorer learning outcomes, compared to the teaching method whereby they individually complete a concept map, the model of completing a concept map through co-operation could further enhance their learning willingness and motivation, and significantly increase their learning outcomes. Therefore, the use of concept maps could encourage students with less learning willingness and motivation to participate in class activities.

Similarly, Hauan and Kolstø (2014) concluded from the existing literature that worksheets with open-ended questions could encourage free exploration in SCV activities. The combination of worksheets containing well-designed questions with appropriate teaching methods can motivate students’ interest in learning. However, Ueckert and Gess-Newsome (2008) pointed out that passive learners would cope with the situation by using the strategy of word-matching: by comparing the questions with the textbook content, the answers from the textbooks are directly copied and pasted onto worksheets. Using worksheets for group work that is integrated with multi-dimensional activities could motivate students aged 13–16 to develop a positive attitude and internal motivation for SCV activities (Stavrova & Urhahne, 2010). Mortensen and Smart (2007) found that among SCV learners aged 8–11, some students preferred open-ended questions while others preferred worksheets consisting of closed-ended questions.

To sum up, if the teaching methods involve interactions between teachers and students and within peer groups, the work can motivate students to learn (Gill, 2007). If the tablet PC, an electronic information and communication medium, is used to help students express their thoughts and exchange concepts among themselves and with their teachers, learning motivation and learning achievement should be further improved. The imparting of knowledge, including phishing knowledge, in a class requires the classification of clear tiers of information situated in clearly conveyed relationships, characteristics possessed by the information in the concept map; in conjunction with the tablet PC’s use as a mechanism for real-time feedback, the concept map should therefore serve as a basis for more effective interactions among and knowledge acquisition by learners. Paper-based worksheets are a frequently administered teaching aid in elementary and secondary schools (Krombaβ & Harms, 2008; Mortensen & Smart, 2007; Stavrova & Urhahne, 2010). However, traditional paper-based worksheets make it difficult for teachers to provide immediate feedback, because they can be returned to students only after teachers have evaluated them (Lee, Luchini, Michael, Norris, & Soloway, 2004). Therefore, this study hypothesizes that learners engaging in in-class learning activities with concept maps will have significantly stronger anti-phishing learning motivation and achievement than learners engaging in in-class learning activities with worksheets.

Materials and methods

Participants

The study participants were 155 eighth- and ninth-grade students (78 male, 77 female) from a junior high school. As of 2012, the school had 15 normal co-ed classes, with approximately 20 students in each class. Purposive sampling was used to classify the participating students into an experimental group (77 students, with an average age of 14.55 years and a standard deviation of 0.50 years) and a control group (78 students, with an average age of 14.48 years and a standard deviation of 0.50 years). Both groups were taught by the same teacher.

Methods and instructional design

Figure 1 presents the quasi-experimental design used in this study. Prior to the class, pretests were conducted using a learning motivation questionnaire (Pintrich, Smith, Garcia, & McKeachie, 1991) and an anti-phishing learning achievement test to gauge students’ anti-phishing learning performance. Subsequently, anti-phishing instruction was used to teach both groups the relevant concepts. Figure 2 shows students answering questions using the tablet PC’s interactive polling tool (Interactive, Feedback-based In-class Teaching system, iFIT). The
anti-phishing instruction was carried out as follows (the experimental and control groups adopted identical anti-phishing teaching methods):

1. The teacher illustrated the severity of global phishing problems with statistics, and presented real-life examples of victims of phishing frauds (from news articles).
2. The teacher defined and explained the concept of phishing (see 2A in Figure 2).
3. The tablet PC polling tool was used to carry out real-time feedback-based learning activities (see 2B, 2C in Figure 2). The following is an example of a polling topic: “What should we do when we suspect that we have been attacked by a phishing site or an online fraudster?”
4. The teacher gauged the students’ current level of understanding of the material and received suggestions from the students regarding the progress of the class via the tablet PCs.
5. The teacher helped students recognize the various forms of phishing, and advised them on how to identify phishing emails and websites.

![Figure 1. Experimental design](image)

After instruction was carried out according to the above steps, the experimental group made use of the concept map for their learning activities; that is, under the teacher’s guidance, the students formed groups in which they drew concept maps of phishing using their tablet PCs (see Figure 3 as an example). The teacher then shared and discussed the individual group’s outcomes (concept maps) with the class as a whole (see 2D in Figure 2). Lastly, the groups supplemented and amended their concept maps on the basis of feedback from the teacher and their classmates.

The control group, on the other hand, was organized into groups and asked to complete worksheets together to help them consolidate their knowledge of the phishing concept. The teacher provided feedback and suggestions based on the students’ answers. To prevent learners from applying matching-word and copy-and-paste strategies when completing the worksheets, this study included both open-ended and closed-ended questions, with the scope of questions ranging from anti-phishing conceptual knowledge and case studies to generating solutions in order to prevent phishing traps. Some sample questions are as follows: “What are the formats of phishing?”
“Please give some examples to illustrate it,” and “What strategies will I adopt in future to prevent falling into phishing traps?”

**Figure 2.** Photos of classes in progress

**Figure 3.** Concept map screenshot on tablet PC

Finally, both the experimental group and the control group completed a posttest on anti-phishing learning motivation and anti-phishing learning achievement, while 20 of the students were also interviewed.

**Variables and statistical analyses**

In this study, the independent variable was “in-class learning activities” (with concept maps or worksheets), the dependent variables were the posttest scores for anti-phishing learning motivation and achievement, and the
covariates were the pretest scores for anti-phishing learning motivation and achievement (see Figure 1). The control variables were as follows: both groups carried out activities in teams, in which the group size was similar (3–4 people in each group), the gender ratio was similar, and the groups had the same teacher. In this study, we used the Statistical Analysis System (SAS) 9.4 and Predictive Analytics Software (PASW) 22 to perform analysis of covariance (ANCOVA). After controlling for the covariates, the independent variables’ effect on the dependent variables was explored. If the slopes of the regressions of the in-class learning activities were not statistically different (no significant ANCOVA interaction term), the data were then tested for mean differences using independent sample t-tests. If the slopes were different, the data were subjected to the Johnson-Neyman procedure (Johnson & Neyman, 1936), which identifies the regions of the covariates where the effect of the independent variable turns from non-significant to significant.

**Instruments**

The anti-phishing learning motivation questionnaire was adapted from the Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich et al., 1991). The scale was a 7-point Likert scale, comprised of 31 questions with three constructs: the expectancy component, the value component, and the affective component. The Cronbach’s α coefficients for the pretest were 0.87, 0.90, and 0.62, respectively, whereas for the overall scale the value was 0.92. All coefficients were above medium reliability (Nunnally, 1978).

The anti-phishing learning achievement test questions were formulated by the researchers and tested by three experts for validity, appropriateness and relevancy; the sequence and difficulty of the questions were adjusted, based on the experts’ suggestions, before finalization. The tests were composed of three sections, respectively covering “forms of phishing,” “the concept of anti-phishing,” and “anti-phishing methods and skills.” There were 10 true/false questions (TFQs) and 8 multiple-choice questions (MCQs). The posttest questions were identical to the pretest questions; however, in order to fully comprehend the students’ learning status, the sequence of questions was revised, as was the order of the MCQ answer options. The Cronbach’s α coefficient of the test questions was 0.73, indicating good reliability (Nunnally, 1978). The discrimination of all the test questions was greater than 0.3, indicating good discrimination as well (Ebel & Frisbie, 1986).

**Results**

**Anti-phishing learning motivation results analysis**

In order to confirm that there was no significant difference in the learning motivation of the experimental group and the control group before the experimental treatment, the independent-samples t-test was used. The variance homogeneity testing did not reach statistical significance (Levene’s test, p = .06), indicating that both groups could be viewed as homogeneous. Further, the average score for initial (pre-test) anti-phishing learning motivation was 4.62 and 4.46 for the control and experimental groups, respectively, where t(153) = 1.23 and p = .22, indicating no significant difference in initial anti-phishing learning motivation between the groups. The average scores of the experimental group and the control group across the three dimensions were 4.73 and 4.53, 4.64 and 4.46, and 3.62 and 3.66 for the expectancy component, the value component, and the affective component, respectively, and the p-values were .07, .09 and .67, respectively, indicating no significant differences in the initial anti-phishing learning motivation between the two groups across these dimensions either. Upon completion of the experimental treatment, both the experimental group and the control group took the anti-phishing learning motivation posttest. The preliminary ANCOVA confirmed a significant pretest score for anti-phishing learning motivation by the in-class learning activities interaction effect (F = 9.58, p < .01). To examine this interaction further and to identify the range of values for which the covariate had a significant effect on the independent variable, we applied the Johnson–Neyman technique (Johnson & Neyman, 1936).

The Johnson–Neyman technique showed two significant regions for the effect of in-class learning activities on the posttest scores for anti-phishing learning motivation: for posttest scores below 3.88 and above 5.19 (see Figure 4). If the learners had pretest scores above 5.19, those who were in the experimental group obtained higher posttest scores than those in the control group (b = 0.29, t = 1.98, p < .05, 95% CI = 0.00 to 0.58). Conversely, the worksheets (control group) led to significantly higher posttest scores for anti-phishing learning motivation than the concept map (experimental group) when the pretest scores were below 3.88 (b = -0.30, t = -1.98, p < .05, 95% CI = -0.60 to 0.00).
Figure 4. Interaction of treatment and pretest scores in predicting posttest scores for anti-phishing learning motivation

Anti-phishing learning achievement results analysis

In order to confirm that there was no significant difference in the initial anti-phishing learning achievement of the experimental group and the control group, both groups completed anti-phishing learning achievement pretest questionnaires; the independent-samples t-test was used to test the pretest learning achievement scores of both groups. The variance homogeneity test did not reach statistical significance (Levene’s test, $p = .08$), indicating that both groups could be viewed as homogeneous. Further, the average pretest scores were 81.44 and 78.93 for the control and experimental groups respectively, where $t(153) = 1.32$ and $p = .19$, indicating no significant difference in the initial anti-phishing learning achievement of the two groups.

Figure 5. Interaction of treatment and pretest scores in predicting posttest scores for anti-phishing learning achievement

ANCOVA was then conducted to compare the effect of the in-class learning activities on the posttest scores for anti-phishing learning achievement, controlling for the pretest scores. A significant interaction effect was found between the pretest scores and the in-class learning activities ($F = 7.32$, $p < .01$), so a modification of the Johnson–Neyman technique (Johnson & Neyman, 1936) was applied. The technique showed one significant region for in-class learning activities on the posttest scores for anti-phishing learning achievement: for pretest
scores below 74.61 (see Figure 5). When the pretest scores were below 74.61, learners in the experimental group obtained significantly higher posttest scores than those in the control group \( (b = -2.92, t = -1.98, p < .05, 95\% CI = -5.84 \) to 0.00). However, when the pretest scores were above 74.61, there was no significant difference between the experimental group and the control group on the posttest scores for anti-phishing learning achievement.

**Discussion**

**The influence of concept map activities in anti-phishing education on learning motivation**

This study found that in the concept map condition, after the anti-phishing instruction, there was significant improvement in the learning motivation when the learners had an initial high learning motivation. Conversely, for learners with below-average initial learning motivation, the level of anti-phishing learning motivation after use of the worksheets significantly increased. A possible reason for this is that concept maps are more interactive and the scaffolding structure prepared by the teachers is less obvious than for the questions on the worksheets. As a result, the use of concept maps requires more proactive participation from teachers and students in areas such as discussion, knowledge construction, and concept connection, and is more suitable for enhancing the interest of students with high learning motivation. Additionally, the introduction of tablet PCs and interactive, in-class activities made the class more interesting, which stimulated the proactive trying and regulation of the students’ learning behavior. Another possible reason for this is that the characteristics of teaching can stimulate students’ learning motivation (Sweller, 1994; Yang et al., 2012). The students with high initial learning motivation participated seriously and conscientiously in this study as compared to their conventional classroom behavior, indicating a positive boost due to the use of the tablet medium (Anderson et al., 2004; Gill, 2007; Mock, 2004). When asked about the reasons for their answers, the students were largely willing to voice their opinions as well as to take the initiative to consider the content of the question and select their answers cautiously.

Teaching using question-and-answer worksheets is a relatively common and familiar education strategy for helping individuals or groups of students reflect on the knowledge they have acquired, which may make them less interesting than a novel method; in contrast, concept maps are a relatively new teaching method, and students are naturally interested in this learning medium; concept maps also provide students with the opportunity to exercise their imagination and creativity, making learning more enjoyable. Observation of students’ participation in class revealed that the students in the experimental group were clearly more proactive than those in the control group, and their group discussion was livelier, indicating that the concept map activities promoted the learning motivation of those who had high initial learning motivation.

This study further strengthens the research results of Liu and Lee (2013), and Ojima (2006). In addition to improving learning motivation and reducing anxiety, our study found that completing concept maps through group work as a teaching method is also applicable to learners with strong learning interest and efficacy. There is also evidence that for learners with less prior knowledge, the teaching method of completing concept maps in groups shows better performance than the individual completion of maps in terms of enhancing their learning motivation and outcomes (Kinchin & Hay, 2005; Liu, 2011). However, the preliminary findings from this study suggest that for learners with lower pre-teaching learning motivation, compared to completing concept maps in groups, completing worksheets in groups could be more effective in terms of enhancing their learning interest. We speculate that this is due to the learning subject being highly specific.

Currently, no textbook specifically focuses on anti-phishing in elementary or secondary education, and in this study, online information searching was prohibited during the worksheet activities to reduce the learners’ frequency of adopting the word-matching strategy (Ueckert & Gess-Newsome, 2008). When the copy-and-paste strategy is not applicable, passive learners would be more attentive in class and would complete closed-ended questions on worksheets. This study not only included questions with model answers, but also included a small number of open-ended questions for learners with low initial motivation to explore and think about. Previous research has already shown that with group work and multi-dimensional learning contents, using worksheets could strengthen learners’ intrinsic motivation (Stavrova & Urhahne, 2010). Our study further revealed that with optimal structuring of questions and applications, the use of worksheets can strongly enhance learners’ motivation, in particular among learners who lack initial learning interest, and using worksheets helps these learners participate in and concentrate more on the learning contents and activities.
The influence of concept map activities on anti-phishing learning achievement

The study found that when the level of prior knowledge was low, the posttest scores of the control group were significantly higher than those of the experimental group. For learners with sufficient prior knowledge, regardless of the type of in-class activities, their post-experimental test scores significantly improved, and both worksheets and concept map teaching were applicable to them. After the teacher verbally taught students about anti-phishing, the in-class learning activities provided them with another opportunity to recall and consolidate what they had learned, so that the concepts would form a complete knowledge architecture for those with sufficient prior knowledge. Therefore, both types of in-class learning activities could help students extract concepts of knowledge and lower the cognitive load of the learning materials (Sweller, 1994). It may be that anti-phishing was not an entirely new or problematically complex concept for them. Therefore, irrespective of the teaching method, they seemed to have been able to master the key concepts during the learning process, and so their test scores naturally improved (Moen & Boersma, 1997). Another possible reason is that the concept maps emphasized the links between concepts, meaning that the students needed to understand every concept presented before they could fully construct an overall framework (Cliburn Jr, 1987). As the cognitive ability of junior high students is still limited, it was a challenging task for them to establish the relationships among concepts, so the concept map activities could not increase their cognitive resources, which were used for building concepts in the learning tasks, resulting in the finding that the experimental group’s posttest performance was not significantly higher than that of the control group. Further, as concept mapping software is still at an immature stage, it was notably difficult to construct and use the concept maps on the tablet PCs. The hindrance posed by the learning interface reduced the learning achievement compared to the worksheets. In the interviews, two students mentioned that the operation of the concept maps on the tablet PCs was time-consuming and did not seem to help much in consolidating the concepts. Therefore, immature concept mapping tools might cause operational difficulties for learners with sufficient prior knowledge. No differences were found in the posttest scores between the experimental group and the control group.

In contrast, for students with insufficient prior knowledge, the worksheet questions were used to stimulate their thinking, and when they were familiar with or clear on the intention of the questions, their learning efforts yielded even better outcomes. Previous research found that students with low prior knowledge and basic skills are better suited for the teaching method of group completion of concept maps, while students with sufficient prior knowledge are more suited to individual completion of concept maps (Kinchin & Hay, 2005; Liu, 2011). However, completing concept maps requires thorough comprehension and advanced skills of making connections (Cliburn Jr, 1987). It is inferred in this study that students with less prior knowledge failed to effectively use their cognitive resources to learn the structure of the concepts and summarize them; however, as worksheets present a clearer knowledge structure that is prepared by teachers, students with less prior knowledge are able to effectively use them to summarize key concepts (Cliburn Jr, 1987). The summary review function of worksheets ensures a learning outcome superior to that of concept maps for these students.

Conclusion and recommendations for future study

The purpose of this study was to investigate the feasibility of the integration of concept maps and tablet PCs in anti-phishing education to enhance students’ learning motivation and learning achievement. The findings show that when learners’ pretest learning motivation was high, the in-class activity with concept maps on tablet PCs significantly increased their posttest learning motivation. On the contrary, for learners with low pretest learning motivation, their posttest learning motivation after the worksheet activity was significantly higher than that of the students in the experimental group. In terms of anti-phishing learning achievement, for learners with low pretest learning achievement, compared to the concept map activity, the use of worksheets could significantly increase their posttest learning achievement. On the other hand, for learners with high pretest learning achievement, there were no significant differences between the control and the experimental groups in their learning outcomes. On this basis, this study proposes that anti-phishing learning motivation and learning achievement can be enhanced if the anti-phishing curriculum is designed to reflect students’ initial learning motivation or prior knowledge and to integrate mature and accessible technological media. These findings should be able to serve as a reference for educators and researchers working on anti-phishing education or on the integration of tablet PCs and/or concept maps in education.

Even though the two types of activities could produce the same learning outcomes for students with more prior knowledge, the use of concept maps is able to increase the motivation of learners interested in anti-phishing issues. The results of our research show that the correlation between pretest motivation and pretest achievement was statistically insignificant ($r = .07, p = .42$). This means that, regardless of the amount of prior knowledge,
learners’ interest in anti-phishing issues varied from person to person. For students with sufficient basic knowledge, although both activities could enhance their posttest scores, teachers should choose different teaching activities according to the students’ motivation levels: if their motivation is above average, concept maps can be used for teaching, while if it is below average, teachers can conduct worksheet activities. For students lacking basic knowledge, if their motivation is below average, worksheets can be used to help them construct knowledge and increase their learning interest in anti-phishing. If their motivation is above average, teachers can employ a more dynamic education strategy, such as integrating both worksheets and concept maps into a tablet PC. This will help the students make use of the worksheets to construct concepts and gain knowledge while simultaneously using concept map blank-filling games as a tool for reviewing lessons or increasing motivation.

However, the present study also shows that there are still some limitations to our ability to integrate such technologies into real-life teaching. Concept maps emphasize the formation and organization of concepts by an individual (or a group). Students often generate new ideas during group discussion and wish to alter a previously developed concept map to reflect them. Conventional concept maps are drawn on paper, which can be easily amended; in contrast, in the case of maps developed on tablet PCs, wrong keys may be pressed and students may need to type in amendments using the on-screen keyboard; therefore, the amendment process may not be as easy as expected for junior high students who are not familiar with tablet PCs, and their learning outcomes may be limited as a result. There are already some user-friendly concept mapping tablet PC applications on the market (such as Total Recall and Idea Sketch). Although the system used in this study possessed some useful capabilities such as uploading and sharing functions, the difficulty of drawing concept maps on the tablet PCs was found to be quite high, as was that of adding/removing or linking concepts. Therefore, this study proposes that future work combine concept maps and tablet PCs in class in the following two ways: (1) draw the concept map on paper and use the camera function of the tablet PC to upload the concept map onto the website; and (2) integrate screen-shot and uploading functions into existing concept map software, then the concept map can be uploaded onto the teacher’s web portal.

Further, in order to encourage students to be more proactive in exploring and acquiring more advanced knowledge, this study proposes the adoption of dynamic teaching methods in anti-phishing education, for instance increasing the interaction between teachers and students (Sheng et al., 2007), in-class tablet-PC games, or group collaboration (Sheng et al., 2007; Yang et al., 2012), so as to make students the host of the classroom to proactively participate in the course, thereby improving the efficacy of anti-phishing education. This inference is only applicable to anti-phishing classes. For courses with different characteristics, such as a large amount of teaching material or subjects with complex knowledge, it is necessary to further explore the effect of multidimensional and dynamic teaching methods on various teaching methods such as encouraging students to actively participate in learning, helping students effectively construct knowledge by using cognitive resources, and improving learning motivation.

Acknowledgements

This research was supported by the Ministry of Science and Technology in Taiwan through Grant numbers MOST 104-2511-S-009-008-MY3 and NSC 101-2511-S-009-010-MY3. The authors would like to thank the students who participated in this study and acknowledge the contributions of Shian-Shyong Tseng, Sunny S. J. Lin, Yu-Yan Lin, and Ting-Yu Lin who supported this research study and provided valuable comments.

References


Effects of a Peer Assessment System Based on a Grid-based Knowledge Classification Approach on Computer Skills Training

Ting-Chia Hsu (A.K.A. Ching-Kun Hsu)
Department of Technology Application and Human Resource Development, National Taiwan Normal University
//ckhsu@ntu.edu.tw

(Submitted April 26, 2015; Revised September 27, 2015; Accepted February 25, 2016)

ABSTRACT
In this study, a peer assessment system using the grid-based knowledge classification approach was developed to improve students’ performance during computer skills training. To evaluate the effectiveness of the proposed approach, an experiment was conducted in a computer skills certification course. The participants were divided into three groups: one that learned with traditional instruction, one given the conventional peer-assessment system, and one using the proposed peer-assessment system. The results showed that the learning achievements of the students using the proposed system were significantly better than those of the other two groups. Therefore, integrating the knowledge engineering approach with the peer-assessment process can benefit students’ learning, and help them attain computer skills certification. The dynamic peer assessment with knowledge classification approach is not only useful, but can also be repeatedly applied to different question sets of the certificate of computer software application.

Keywords
Peer assessment, Grid-based knowledge classification, Knowledge engineering

Introduction

Presently, most vocational students have significant experience using computers, and endeavor to pass the certification of the office word processing software application in their first year of senior vocational school in Taiwan. The skill evaluation center of the workforce development agency of the Council of Labor Affairs (i.e., CLA) conducts such testing. San et al. (2006) indicated that the licensing rate is an important index for measuring labor quality; moreover, those who pass the licensing tests are deemed to be more professionally qualified and competitive in the labor market (San et al., 2006). There is no implementation subject which every vocational school student has to learn besides the certification of computer software application, because fostering such competence is one of the essential employability skills in the workplace. In other words, there is a particular population with an urgent need to learn this topic. The certification includes two stages. The first assesses the students’ academic knowledge, while the second assesses their practical computing skills with real tasks which must be completed within two hours using the functions of the application software for managing, editing, and positioning the required fonts, tables, paragraphs, format styles, word art, visual material, columns, and layout size. After passing both the academic and practical tasks, CLA will certify the students.

Knowing how to operate software does not necessarily mean one can correctly use it for practical applications. Learning to apply the available information and software to solve problems is vital for vocational high school students (Valtonen, Hacklin, Dillon, Vesisenaaho, Kukkonen, & Hietanen, 2012; Chang, Tseng, & Lou, 2012). Scholars have noted that vocational students have relatively low academic achievement and less positive attitudes toward learning, resulting in a lack of higher-order thinking (Lee, Shen, & Tsai, 2010; Yang, 2015). Yang (2015) warned that low engagement and outdated approaches to instruction are the two main challenges of vocational education. Repeated operation without supporting students’ thinking may result in a lack of problem-solving capability in the workplace (Yang, 2015). Therefore, it is necessary to provide these students with higher-order thinking learning activities. Simply learning a software’s functionality in class does not elicit the students’ full potential because they may not be able to correctly apply, analyze or evaluate the functions they have learned to solve the problems or complete the tasks they have been assigned. To deal with this phenomenon, this study considers a peer assessment approach which lets the students learn not only from the teacher’s lessons but also by finding their peers’ implementation errors (Race, 2001; Hsia, Huang, & Hwang, 2015). Peer assessment is one of the higher-order thinking learning activities (Lorente-Catalán & Kirk, 2016). Moreover, performing the active role of a reviewer can have a positive impact on one’s own learning (Chinn, 2005). Therefore, a peer assessment system with a grid-based knowledge classification approach based on knowledge engineering was developed.

Scholars have indicated that the peer assessment process is challenging for students because they have difficulty knowing what and how to explain their evaluation; however, providing them with specific assistance and
conducting knowledge acquisition are two main strategies for overcoming these problems (Lawrence & Zollinger, 2015). Therefore, the current study hypothesized that using a repertory grid method, one of the knowledge acquisition approaches in knowledge engineering, would help the students conduct peer assessment and learn from the process. Knowledge engineering refers to the process of developing expert systems which are artificial intelligence programs capable of imitating human decision-making processes on the basis of knowledge derived from domain experts. Among the existing approaches to knowledge acquisition, the repertory grid method, originating from the personal construct theory (Kelly, 1955), provides valuable assistance in the organization and presentation of domain experience and knowledge (Hsu, Hwang, & Chang, 2010; 2013). In a standard computer skills certification test, students need to identify and implement the functions required for the assigned tasks in a limited amount of time; therefore, they must completely understand the evaluation criteria of the computer system plus its individual functions. Consequently, referring to the work of others and receiving suggestions from others during the learning process is likely to be helpful, further supporting the use of peer-assessment strategies in school settings (Reid, 2000).

Although numerous studies have reported the effectiveness of peer assessment, some researchers have indicated that students might not benefit from it for several reasons (Chang & Tseng, 2011). For example, the assessment criteria and forms could be difficult to understand, causing them to spend much time deciphering the assessment items (Chang, Tseng, & Lou, 2012). Moreover, it could be difficult for them to evaluate and compare peers’ work and summarize their findings without assistance or boundaries (Gijbels et al., 2005). Meanwhile, the scholars emphasized that finding ways to support and improve the question- or criteria-posing ability and performance of lower-achieving students is an important issue for instructors (Yu, Liu, & Chan, 2005). Additionally, there has been a call for support mechanisms to help make question-posing and peer assessment more manageable (Yu, Liu, & Chan, 2005). Because knowledge engineering helps people organize their acquired domain knowledge or expertise, it has been identified as a potential approach to help students organize what they have learned and experienced during their coursework (Liu & Lee, 2005). Therefore, in this study, the grid-based knowledge classification approach was used to develop a peer-assessment system for helping students evaluate their peers’ performance. To appraise the effectiveness of this innovative approach, an experiment was conducted in a computer skills certification course at a vocational high school with three groups learning through different experimental treatments with or without a peer assessment tool. It was not certain that peer assessment would benefit the majority of students who have relatively low achievement without or with a supportive design (i.e., the grid-based knowledge classification approach). Therefore, three research questions were proposed as follows.

- Do the students make progress week by week after the experimental treatments?
- Do the students using the proposed approach outperform those using conventional approaches?
- Do the low-achievement students in the experimental group gain more benefits than those in the two control groups?

**Literature review**

**Peer assessment**

Falchikov (1995) defined peer assessment as mutual assessment of students in the same grade. Topping (1998) later proposed that it is a process by which students with similar levels of knowledge and learning backgrounds alternately assume the role of reviewers. By doing so, they can judge each other’s learning outcomes. Topping further indicated that integrating technological advances into assessment has the potential to assist in carrying out peer assessments (Topping, 1998). Peer assessment was also conducted as part of the overall summative course assessment in an engineering education course (Hersam, Luna, & Light, 2004). It has been found to have high reliability and effectiveness (Ohland et al., 2005).

In the past decade, various web-based peer assessment systems have been applied in numerous domains, including computer science, science projects, and webpage design (Chen, 2010; Lin, Liu, & Yuan, 2001; Sung, Chang, Chiou, & Hou, 2005). Several studies have reported the benefits of peer assessment (Boud, Cohen, & Sampson, 2014; Kaufman & Schunn, 2011; Strijbos & Sluijsmans, 2010). For example, Alexander et al. (2008) compared the differences between traditional and peer assessment models and found that engaging students in peer-assessment activities improved their learning outcomes. Van Zundert, Sluijsmans, and van Merriënboer (2010) showed that peer assessment benefited the development of students’ domain-specific skills via the evaluation, reflection, and revision cycle. Recently, one study showed that there was a positive correlation between the ratings given by teachers and peers, although the two groups interpreted the rubric criteria differently (De Grez, Valcke, & Roozen, 2012).
A peer assessment system entails consideration of peers’ accomplishments, and requires students to use higher-order thinking skills such as analyzing, solving problems and evaluating phenomena rather than memorizing facts (Bloom, Englehart, Hill, & Krathwohl, 1956; Barak & Rafaeli, 2004; Sluijsmans et al., 2001; Zohar & Dori, 2003). Peer assessment is a higher-order thinking learning activity and was hypothesized to be a proper strategy for students to learn from checking others’ work. However, researchers have indicated that without careful design of the assessment criteria and forms, it could be very difficult for students to understand, especially considering its purpose as a tool to enhance learning through clear peer-to-peer communication; hence peer assessment could in fact act as a learning barrier for students (Chang et al., 2012). Moreover, without assistance, students might be unable to organize and analyze their assessment results, implying the necessity and importance of developing effective peer-assessment strategies (Gijbels et al., 2005). This is why the current study proposes a dynamic repertory grid approach for a peer assessment system used in a computer skills class, allowing student reviewers to select personalized assessment criteria from a bank of items, causing them to reflect on the purpose of the assignment being reviewed and how to best characterize the performance of the student completing the assignment, as well as providing ample feedback for the students.

The grid-based knowledge classification approach in e-learning

In recent decades, researchers have proposed various knowledge engineering methods to help knowledge-based system developers acquire and organize knowledge from domain experts (Keynan, Ben-Zvi Assaraf, & Goldman, 2014; Siau, Tan, & Sheng, 2010). The repertory grid method proposed by Kelly (1955) is one of the most frequently adopted knowledge acquisition methods owing to its simplicity and clear structure (Malmström, Johansson, & Wincent, 2015; Kumar & Natarajan, 2010). Kelly (1955) suggested that people can create explanations of phenomena from their experiences, which can help them develop their understanding of underlying concepts and can be used as the basis of future judgments.

A repertory grid consists of three major components: elements, constructs, and ratings (Fransella, Bell, & Bannister, 2004). It can be viewed as a matrix with columns containing element labels and rows containing construct labels. Elements can be decisions to be made, objects requiring classification, or concepts to be learned (Chu et al., 2010). Constructs are features used to describe similarities or differences among elements. Each construct describes a trait and its opposite. A 5-point scale rating mechanism is commonly used to represent relationships between elements and constructs, where a score of 1 indicates that the element is very likely to have the trait, 2 indicates that it may have the trait, 3 corresponds to responses of “unknown” or “no relevance,” 4 indicates that the element may possess the opposite trait, and 5 indicates that it very likely possesses the opposite trait (Fallman & Waterworth, 2010; Hsu et al., 2010; 2013).

Repertory grids have been employed to help researchers collect and organize domain knowledge for the development of expert systems, which, as previously defined, are artificial intelligence programs designed to simulate expert reasoning on the basis of knowledge elicited from domain experts (Keynan et al., 2014). Repertory grids have not only been used to integrate opinions from various experts to accumulate important data and preserve it in a database (Hoffman, Shadbolt, Burton, & Klein, 1995; Abdul-Gader & Kozar, 1990), but are also seen as an important mind-tool for e-learning (Hwang, Hung, Chen, & Liu, 2014). In our study, repertory grids are a knowledge classification approach for helping students organize their assessment results and illustrate their review judgements. This approach is introduced in the following section.

Method

Participants

A total of 224 tenth graders whose average age was 16, from six classes, participated in this study. They studied computer software applications for two hours a week in a senior vocational school. This was the first semester they had taken the computer software application certificate course. They were therefore beginners preparing for the Council of Labor Affairs certification. One class was assigned to be the experimental group, while the other two classes were assigned to be control groups 1 and 2. The experimental group did peer evaluation with the online peer assessment system based on the grid-based knowledge classification approach, while control group 2 used the conventional peer-assessment system. Every student in these two groups acted as a practitioner and a reviewer, and so had to assess one set of implementation work done by another student. Control group 1 did not perform peer assessment. All of the work of the groups was evaluated by the teacher each week. The six classes
were taught by the same instructor who did the certification evaluation with another expert. Both of these teachers had more than ten years’ vocational teaching experience.

The conventional peer-assessment system

Control group 2 was issued the assessment form used by the CLA as their peer-assessment form. Figure 1 shows the peer-assessment criteria used in the conventional CLA peer-assessment approach. The rows in the assessment form are fixed, preventing students from revising them. Thus, the peer-reviewers were only allowed to fill in the number of errors without any input into designating the assessment criteria.

The computer application certificate includes the most popular Office suites, information processing, word-processing, and document formatting. Some users may think they know how to format a document; however, they may not use appropriate, correct or efficient functions. Are they aware of the mistakes they make during the test? This study conducted experiments to compare the effectiveness of the peer assessment systems and traditional instruction.

A peer assessment system based on a grid-based knowledge classification approach

An alternative system which is a repertory grid-based peer assessment tool, was developed for the experimental group. Two instructors specializing in computer software certification were invited to design an assessment form based on the repertory grid approach. Table 1 presents an illustrative example of a repertory grid in which the 10 elements (i.e., $P_1$, $P_2$, ..., $P_{10}$) represent 10 paragraphs in an article edited by a student, and the 19 categories ($C_1$, $C_2$, ..., $C_{19}$) represent a basic structure of peer-assessment assistance. In each category (e.g., paragraph format), a set of sub-items (e.g., indent in the first line of the paragraph and space between the lines in the paragraph) could be selected for assessing the paragraphs via a drop-down menu. The students were required to identify the relevance of each sub-item to the assignment they were assessing, and choose the items themselves when assessing their peers’ work. In this way, a partially self-constructed assessment form was used during the assessment process. In this study, we attempted to provide the students with assistance when they conducted the higher-order thinking learning activity of peer assessment. Engaging them in identifying the importance of the sub-items and constructing their own assessment forms was intended to support their higher-order thinking, as suggested by several researchers (Barak & Dori, 2009; Van den Berg, 2004).

In the repertory grid-based assessment form, a rating of 1 means the student correctly employed or developed the function, showing that he/she had fully developed the computer skill described in the construct (e.g., data processing skill); 2 means that he/she is equipped with a sufficient but incomplete computer skill; 3 means that he/she has learned part of the skill; 4 means that he/she seems to have failed to use the appropriate computer skill to fulfill the requirements; 5 means that he/she has completely failed to learn the skill. Note that the peer-
assessment process was conducted anonymously. Figure 2 presents an illustration of the repertory grid assessment form. Using the repertory grid developed for this study, the development procedure of the peer assessment activity is given as follows.

Table 1. Illustrative example of a repertory grid

<table>
<thead>
<tr>
<th>Positive constructs</th>
<th>Elements</th>
<th>Opposite constructs</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1. Good paragraph format (e.g., first line of the paragraph is indented by 2 characters)</td>
<td>P1 1 P2 1 P3 5 ... P10 5</td>
<td>C1’. Poor paragraph format (e.g., first line of the paragraph is not indented by 2 characters)</td>
</tr>
<tr>
<td>C2. Good font format (e.g., using the standard font, such as Times New Roman)</td>
<td>5 1 2 ... 4</td>
<td>C2’. Poor font format (e.g., using various fonts or incorrect fonts)</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>C19. Good list and bullet format (e.g., the list and bullet format are used when necessary)</td>
<td>5 5 1 ... 1</td>
<td>C19’. Poor list and bullet format (e.g., abusing the list and bullet format)</td>
</tr>
</tbody>
</table>

![Figure 2. Experimental group assessment form](image)

Step 1: Choose sub-items from the drop-down menu in each available category on the basis of the requirements of the certificate questions. After choosing one sub-item, a construct for evaluating each paragraph is produced, with the construct’s description and its opposite shown on the left and right sides of the repertory grid, respectively. It is shown from (a) to (e) in Figure 2.

Step 2: Check the criteria of the learning task and select additional sub-items from the categories if needed.

Step 3: Evaluate each paragraph on the basis of the selected sub-items by filling in the rating, ranging from 1 to 5 in order of best to worst, for each <item, sub-item> entry.

Step 4: Provide qualitative comments at the end of the assessment form (shown as (f) in Figure 2).

Step 5: Submit the assessment results (shown as (g) in Figure 2).
Experimental design

Students in control group 2 and the experimental group participated in peer assessments every week, whereas students in control group 1 were assessed only by the instructor. The experiment was conducted over a period of six weeks. Each week the students did a certificate item set and then handed over their implementation. The first week named week 0 was the pretest, then the results were obtained in four weekly stages named week 1 to week 4, while the last week named week 5 was the posttest. The students in each class were further grouped into three achievement levels according to their achievement on the pretest.

Two independent variables were included in the experiment design: pretest achievement and experiment treatments (control and experimental groups). The covariate was the pretest scores representing the students’ prior achievements. The experimental procedure is outlined in Figure 3. The dependent variable was the students’ learning outcomes.

![Figure 3. Experimental procedure](image)

Data collection

The government certificate database includes 15 data processing projects (tests), all of which have been proved to have equal degrees of difficulty and discrimination. The 15 tests are comparable, so every student taking the official certification test is randomly assigned only one. The 15 tests involve similar required functions from the application software, although their implementation appearances are different. The six tests in this study including the pre-test at the beginning, the 4 tests during the following 4 weeks, and the post-test in the final week, were randomly selected from the 15 projects (tests).

Results

Weekly improvement

There was no significant difference among the three groups in the pre-test \( F = 0.229, p > .05 \). The experimental results show that the integration of peer assessment and knowledge engineering guided the students to focus on constructing solutions from the concepts they had learned and integrating their computer software skills to complete the tasks they were assigned. The achievements of the participants in each group are presented in Figure 4. The students in the first control group did not make progress in some weeks, and their performance in week 3 was not significantly better than in the pretest. The second control group made slow progress, indicating that peer assessment with an inappropriate assessment tool does not necessarily produce ideal results.

A 3 x 5 mixed ANCOVA with experimental treatments as a between-subjects variable, time-point (weeks 1-4 and posttest) as a within-subjects variable, and the pretest as the covariate was employed. The analysis allowed concise description of whether there was an effect of the experimental conditions across the time-points, or possibly interaction between the two. The results showed that there was no interaction between the two variables \( F = 0.919, p > .05 \). The main effects of the tests across different weeks had significant differences \( F = 10.775^{***}, p < .001 \). In other words, there was remarkable improvement among the weeks based on the evidence from the within-subjects effect test. After multiple comparisons, it was found that the effectiveness in the second
and fourth week was significantly higher than that in the first week when the pre-test was the covariance. The means and standard deviations can also be seen in Table 2. Overall, the post-tests of each group were all remarkably higher than their pre-test.

Multiple comparisons among groups

Based on the results of the 3 x 5 mixed ANCOVA, the main effect of the experimental treatment was found to be significant \( (F = 11.552^{***}; p < .001) \). In other words, the experimental group performed significantly better overall than control groups 1 and 2 after the paired comparisons. To explore the weekly progress and the difference among groups, the means, standard deviations and adjusted means are shown in Table 2.

From the tests of the between-group effects, listed in Table 2, there are significant differences in the post-test results of the three groups \( (F = 7.90^{***}; p < .001) \). This made it necessary to conduct multiple comparisons, which revealed that the learning results of the experimental group were significantly better than those of the two control groups. However, no significant difference was observed between the learning achievements of the two control groups. In other words, the conventional peer assessment strategy did not result in better learning achievements in comparison with the traditional teaching method. In week 3, both the experimental group and control group 2 significantly outperformed control group 1 \( (F = 6.34^{*}; p < .05) \). Moreover, it was found that the learning achievements of control group 1 approached those of control group 2 in the week 4 and post-test, shown as Table 2.

Table 2. The weekly adjusted mean after ANCOVA analysis

<table>
<thead>
<tr>
<th>Time</th>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Adjusted mean</th>
<th>SE</th>
<th>F</th>
<th>Pairwise comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1</td>
<td>Control group 1 (a)</td>
<td>76</td>
<td>62.92</td>
<td>30.28</td>
<td>63.72</td>
<td>2.78</td>
<td>2.34</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control group 2 (b)</td>
<td>69</td>
<td>62.13</td>
<td>30.02</td>
<td>62.64</td>
<td>2.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Experimental group (c)</td>
<td>79</td>
<td>71.71</td>
<td>27.43</td>
<td>70.50</td>
<td>2.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 2</td>
<td>Control group 1 (a)</td>
<td>76</td>
<td>70.79</td>
<td>30.71</td>
<td>71.37</td>
<td>2.60</td>
<td>1.15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Control group 2 (b)</td>
<td>69</td>
<td>71.25</td>
<td>27.42</td>
<td>71.62</td>
<td>2.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Experimental group (c)</td>
<td>79</td>
<td>77.16</td>
<td>16.85</td>
<td>76.28</td>
<td>2.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 3</td>
<td>Control group 1 (a)</td>
<td>76</td>
<td>60.67</td>
<td>31.03</td>
<td>61.16</td>
<td>2.82</td>
<td>6.34*</td>
<td>b &gt; a*</td>
</tr>
<tr>
<td></td>
<td>Control group 2 (b)</td>
<td>69</td>
<td>70.81</td>
<td>23.22</td>
<td>71.12</td>
<td>2.96</td>
<td></td>
<td>c &gt; a*</td>
</tr>
<tr>
<td></td>
<td>Experimental group (c)</td>
<td>79</td>
<td>75.56</td>
<td>24.23</td>
<td>74.82</td>
<td>2.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 4</td>
<td>Control group 1 (a)</td>
<td>76</td>
<td>74.08</td>
<td>20.54</td>
<td>74.54</td>
<td>2.07</td>
<td>2.13</td>
<td>c &gt; b*</td>
</tr>
<tr>
<td></td>
<td>Control group 2 (b)</td>
<td>69</td>
<td>72.64</td>
<td>23.66</td>
<td>72.94</td>
<td>2.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Experimental group (c)</td>
<td>79</td>
<td>79.53</td>
<td>16.88</td>
<td>78.83</td>
<td>2.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post-test</td>
<td>Control group 1 (a)</td>
<td>76</td>
<td>72.72</td>
<td>29.53</td>
<td>73.20</td>
<td>2.36</td>
<td>7.90***</td>
<td>c &gt; a*</td>
</tr>
<tr>
<td></td>
<td>Control group 2 (b)</td>
<td>69</td>
<td>72.20</td>
<td>22.84</td>
<td>72.51</td>
<td>2.47</td>
<td></td>
<td>c &gt; b*</td>
</tr>
<tr>
<td></td>
<td>Experimental group (c)</td>
<td>79</td>
<td>85.01</td>
<td>13.02</td>
<td>84.28</td>
<td>2.31</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .05; **p < .001.
The improvements of the low-achievement students

From the above-mentioned results, there was no significant difference among the three groups in the pre-test ($F = 0.229$, $p > .05$), but there was a remarkable main effect of the tests across the weeks. Overall, the students made progress in the post-test compared with the pre-test. The benefits which the low-achievement group gained with the assistance of the Grid-based Knowledge Classification Approach for higher-order thinking activities were further explored. Prior to the experiment, the students in each group were divided into three achievement levels on the basis of their pretest scores. In addition to no significant difference being found among the three groups in the pre-test overall ($F = 0.229$, $p > .05$), the low-achieving students ($F = 1.799$, $p > .05$) also displayed no significant difference across the three groups. After the experimental treatments, the low-achievement students in the experimental group gained significantly more than the low-achievement students in control group 1 did ($F = 4.696^*$, $p < .05$). However, there was no significant difference between the improvement of the low-achievement students in control groups 1 and 2. The low-achieving students in the experimental group showed the greatest progress over the course of the learning activity.

<table>
<thead>
<tr>
<th>Tests</th>
<th>N</th>
<th>Mean of gain scores (Posttest-Pretest)</th>
<th>$F$</th>
<th>Comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control group 1 (a)</td>
<td>26</td>
<td>22.96</td>
<td>4.696*</td>
<td>c &gt; a</td>
</tr>
<tr>
<td>Control group 2 (b)</td>
<td>23</td>
<td>29.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental group (c)</td>
<td>26</td>
<td>46.65</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. $p < .05$.

Discussion and conclusions

The results of this study indicated that the assessment form used by the second control group was unable to provide the same benefits as the grid-based knowledge classification approach used in the experimental group. Based on the study of Tseng and Tsai (2007), who conducted an experiment in a computer course in a senior high school and found that the students made significant progress after three rounds of peer-assessment for the same project, it was expected that every group in this study would make similar progress after several iterations of peer assessment. However, it was confirmed that the students who learned with the peer assessment system in control group 2 did not improve as much as those who learned with the peer-assessment approach in the experimental group. The conventional assessment system is the traditional tool which did not provide any further assistance. Both teachers and students had used this traditional assessment form before. After comparing the grid-based knowledge classification approach with the conventional approach, it was found that the former supported the students in conducting higher-order thinking learning activities and in making greater advances as a result of the peer assessment. It is reasonable to infer that the assessment form used in the conventional assessment system was too straightforward, because little thought is required to fill it out; thus, the students did not need to reflect much on what they had learned, resulting in the limited development of their higher-order thinking ability. Despite the fact that the repertory grid can produce peer assessment forms that include partially self-selected items, the students were shown to learn a great deal through construct selection for the evaluation of others’ work. From the evidence of the system logs, it was found that the students in the experimental group selected almost all the possible assessment items (mean = 122) from the selection pools in the top-down menu (i.e., totally 143 choices from 19 categories) in the first week. However, from the third week, they had more confidence and selected only necessary assessment items based on the task requirements. Therefore, the number of assessment items selected stabilized at around half that of the first week (mean = 67). It is thus inferred that the students in the experimental group developed related evaluation thinking ability and learned from their experiences as well as from the mistakes of others week by week. In contrast, the students in the first control group had to accumulate experience by themselves.

The students in the experimental group did not necessarily have the ability to appropriately use the assessment form. Due to the design of the experimental system, when the students in the experimental group were unsure of their judgments, they were able to choose “may” (i.e., “2”) or “may not” (i.e., “4”) as a reference for their peers. As the data remain in the system, the teacher was able to identify the students’ weaknesses. As the system used by the experimental group left indistinct choices for the learners, those choices would remind them of the areas where their peer reviewer was not certain, and the student-reviewers would have to strengthen their own knowledge and skills related to that target. Consequently, the grid-based knowledge classification approach in this study provided the students with a structured and reflective peer assessment system. Although considerable research on the application of repertory grids in peer assessment has been performed (Liu & Lee, 2005), this
study proposes a more flexible instrument than those previously presented. The repertory grid used in this study differs from prior instruments (Liu & Lee, 2005; Liu & Tsai, 2005; Tsai, Lin & Yuan, 2002) which offered a fixed or dynamic grid without classification assistance. The novel contribution of this study is the use of a dynamic repertory grid assessment form to aid learning activities which prepare students for the computer software application certificate. The term “dynamic” here means the students independently chose evaluation items from the drop-down menu under each fixed category. However, the limitation of this study is that the assessment content was designed for one particular certificate, although it is the one which most vocational school students aim to acquire. Another innovation of this study is that it takes the students’ ability into consideration. The instructors designed 19 fixed categories in advance, and offered a pool of several sub-items in the drop-down menu in each category on the basis of the function of the computer application software. The students selected evaluation items provided by the experts from the drop-down menu, which caused them to think analytically about how to best evaluate another’s work and better comprehend the requirements of the set problem to be able to choose proper evaluation items. Boundaries for their choices are thus provided; otherwise, novices may have difficulty establishing the evaluation items independently.

Students are guided to reflect deeply on their learning when they add an assessment item (i.e., a construct in the repertory grid), revise or delete a construct, and compare different portions of a task to assess their peer’s implementation. Previous research indicates that awareness and reflection can help develop students’ metacognition, enhancing their learning and creative abilities (Burleson, 2005). Therefore, in the experimental group, the students were able to construct assessment items to enhance their knowledge and experience reflective thinking through peer assessment. Scholars have remarked that technology can be used as a kind of mind-tool to support students in experiencing reflective thinking that is necessary for meaningful learning (Jonassen & Carr, 2000). The experimental approach left an irresolute space for the students, and the indistinct choices caught the attention of both the students and the student-reviewers. Peer assessment in the experimental group promoted the students’ self-awareness and active role when they selected assessment items specific to the evaluation of each peer’s project that they reviewed, and used these items to evaluate their peers’ implementation, enhancing their reflective thinking through this semi-self-constructed assessment. This research not only helped the students prepare for the certification of computer software application, but also cultivated their higher-order thinking skills (Forehand, 2010) which may be transferred to other implementation practices in the future (Yang, 2015).

The current results show that the achievements of the experimental group were significantly better than those of the two control groups. The low-achieving students in the experimental group benefited the most in this study and achieved the same target as that of a previous study (Yu, Liu, & Chan, 2005). The peer assessment model could be repeatedly used with many question sets like the projects of the computer software application certificate, which differs from the previous experiment which only used peer assessment for one project in the computer course (Tseng & Tsai, 2007). Collaborating with peers may achieve a level which studying alone cannot reach; this improvement is often referred to as the Zone of Proximal Development (ZPD) (Hung & Chen, 2001). Therefore, future studies could extend these methods to include collaborative peer assessment. When students are grouped heterogeneously to use the peer assessment system, they would have a partner to negotiate and discuss which assessment items to choose based on the requirements of the task when evaluating others. It is worth exploring whether such collaborative peer assessment would be more effective.

In the current study, the repeated iteration of doing peer assessment for one task per week accumulated and transferred the skills which the students learned and reflected on to the following tasks. We are planning to apply the dynamic repertory grid method to the peer-assessment activities of other subjects. We also hope to use serial analysis to code and analyze the collected qualitative data (i.e., the students’ comments to peers) to determine whether different peer assessment patterns can be shown to cause different levels of thinking or diverse feedback styles in the innovative assessment mechanism proposed by this study. Finally, allowing students to take part in discussing and setting the criteria for peer assessment may be a feasible way to enhance their learning in computer-related courses in the future (Lai & Hwang, 2015).

**Acknowledgements**

This study is supported in part by the Ministry of Science and Technology in Taiwan under contract numbers: MOST 103-2628-S-003-003-MY2 and MOST 104-2511-S-003-034.
References


Design and Implementation of Technology Enabled Affective Learning Using Fusion of Bio-physical and Facial Expression

Arindam Ray* and Amlan Chakrabarti

1Central Institute of Technology, Chhattisgarh Swami Vivekanand Technical University Bhilai, Chhattisgarh, India // 2A K Choudhury School of Information Technology, University of Calcutta, Kolkata, West Bengal, India

*Corresponding author

(Submitted May 15, 2015; Revised November 24, 2015; Accepted February 1, 2016)

ABSTRACT

Technology Enabled Learning is a cognitive, constructive, systematic, collaborative learning procedure, which transforms teaching-learning pedagogy where role of emotion is very often neglected. Emotion plays significant role in the cognitive process of human being, so the transformation is incomplete without capturing the learner’s emotional state. This paper, proposes a new Affective Computing Module in E-Learning system that focuses on capturing of emotional aspects, managing learning activities, timing and their reflection on the overall learning process. Human Computer Interaction aims at recognizing the learner’s emotional state in order to provide a horizontal interface between human and computer. We have adopted bio-physical (Heart Rate, Skin Conductance and Blood Volume Pressure) and facial expression methods, as these can be made both in-obtrusive and robust against number of environmental conditions, to extract affective state of the learner. This study explores how emotion evolves during the learning process and how emotion feedback is used to improve learning experiences. We propose a learner emotion detection and automated lesson selection model, where two emotion attributes have been fused and achieved an efficient model for Affective E-Learning System, which obtained the sound classification rate. The system can select the learning pedagogy of the learner from detected emotion through neuro-fuzzy logic. Our research shows reasonable results in comparison with the existing systems. The result proves that the recognizer system is efficient.

Keywords

E-Learning, Affective learning, Emotion, Bio-physical signal, Facial signal, Fusion

Introduction

In the age of Internet, the tutoring systems have transformed very rapidly. We have given shape of an organization which is proposed to play a complementary role in education by maintaining quality standard at all possible levels using all possible formats of e-education stressing on e-tutoring. Remembering their school and university years, very many people will agree that their emotional state at a lesson or at a test influenced the result of the process. On the scientific level, a fundamental contribution to clarifying the role of positive emotions of the students in the success of learning/teaching process was done by the System of Emotional-Imaginative Teaching, or the EIT-system (Fomichov, & Fomichova, 1997; 2006; Fomichova, & Fomichov, 1996; 2000). For the best tutoring, learner’s current states need to be identified where the key solution is learner’s emotion recognition. Without addressing of emotion feedback, tutoring system is imperfect. Emotions can have enormous effects on learning and play a vital role in decision making; Inclusion of emotion is must in e-education rather e-tutoring system.

The effectiveness of E-learning (Shen, Wang, & Shen, 2009) also depends on establishing two-way communication between mentors and learners, and among learners themselves. Our research objective is to incorporate new paradigm, that is, learner’s emotion capturing and inclusion in conventional E-Learning System, such emotion based E-Learning system is called Affective E-Learning (Sandanayake, Madurapperuma, & Dias, 2011), to get more effective results. In this context learning is viewed as an active, subjective and constructive activity where the learner is at the center of the learning process (Kordaki, 2004). The major changes in our understanding of the nature of learning imply Technology Enabled Blended Learning (Webb, 2014). The objective of our work is to design a complete delivery model where instant learner’s feedback can be captured at the client site through bio-physical and facial expression, which gets communicated to the e-mentor end. This helps to achieve efficient automatic delivery of lesson according to the learner’s need enhancing the productivity and efficiency of E-Learning. For capturing bio-physical signals we have used three attributes Heart Rate (HR), Skin Conductance (SC) and Blood Volume Pressure (BVP) directly from the learner using ProComp5 device and is passed through the DTREG simulator with emotion-detection function and followed by the generation of the confusion matrix. Concurrently, we adopted the direct capturing of facial expressions using spot detection technique using Webcam from learner and generated the confusion matrix. Finally the two confusion matrices
are fused at the decision level and we obtained the resultant confusion matrix. The result is passed through an automatic lesson detection algorithm to detect the lesson as per the learner’s need. Designing of such automated system is a very challenging task, as all the detections have to be online at the learner end (client site).

Affective e-learning model

In the section we discuss the complete model of our proposed Affective Learning methodology. The following model (Figure 1) shows the complete learning process where the affective computing module is connected with traditional E-Learning system 1.0 through the evaluation module. Our new contribution having the fusion and emotion recognition module makes the whole system more adequate and advanced. Figure 1 shows the complete Affective E-Learning model where we can receive learner emotions with four different attributes (Facial, Speech, Gesture, and Bio-physical Signals). We have considered two emotion-attributes viz. Facial Images captured through spot detection technique and Bio-physical Signals captured using Heart Rate, Skin Conductance and Blood Volume Pressure sensing techniques. Images are captured through QHM495LM-3207 web camera placed in the front of the learner and ProComp5 machine is used to sense Heart Rate, Skin Conductance and Blood Volume Pressure. After receiving the signals from the learner, we fuse them to recognize a specific emotion out of six basic categories of emotion viz. Happiness, Sadness, Surprise, Anger, Disgust, Fear, using multiclass SVM classifier. Based on the emotion detected from the learner, the system automatically detects the delivery-lesson using Fuzzy Logic & Artificial Neural Network. Finally the signal forwarded to Traditional E-Learning System 1.0 for content searching and delivery of the selected lesson. In this paper, we deeply focus on emotion retrieval through Facial and Bio-physical signals.'
Related work

In (Fomichov, & Fomichova, 2012), the authors grounded the necessity of much earlier socialization of children in the Internet age, of making children and adolescents be aware of possible social consequences of their misuse of information and communication technologies. Using special graphs, the authors stated a method of transforming step by step the cognitive-emotional sphere of the learners. The goal is to form the cognitive-emotional subspace of moral values and social responsibility. This paper proposes also an original algorithm of transforming the negative emotions caused by the messages received from social networks into the positive emotions. In (Fomichov & Fomichova, 2014), authors constructively presented the process of education when the values of the student act like a lighthouse for the teacher at the moment of presenting material and arranging the process of education, the process of acquiring knowledge.

The educational methods stated in two mentioned papers are the components of the System of Emotional-Imaginative Teaching. In (Picard, 2003), authors have identified the need of emotions in E-Learning and they showed its application, but no formal mechanism for measuring the effective state of emotion has been explained. Authors (Qin, Zheng, & Li, 2014) focused strictly on negative emotions of e-learners, integrating emotion regulation theories with recommender technique, and present the study of learner-oriented negative emotion compensation based on recommender users and music in the paper. Authors, in (Ho, Nakamori, Ho, & Lim, 2016) have projected that blended learning is effective, flexible, interactive and explained the tool used in it. But they have not focused on emotion recognition concepts. In (Kollias & Karpouzis, 2005), authors have recognized that conventional emotion recognition methods using facial images. Authors in (Kollias & Karpouzis, 2005; Kessous, Castellano, & Caridakis, 2010) didn’t identify how bio-physical signal would be accepted for emotion sensing. In (Kessous, Castellano, & Caridakis, 2010), authors have focused on emotion recognition through Electroencephalography (EEG). But EEG is not recommended at client site in E-Learning platform. According to (Park, Ryu, Sohn, & Cho, 2007) emotional intelligence can help to enhance learners’ positive attitude but focused on learning aspect only. In (Lei & Morrow, 2010), author’s empirical study focused on different learning pedagogy in E-Learning and tried to establish a virtual class room where an emotion sharing between learner and teacher is desirable. In (Ferreira, 2012), the author has proposed several tools for pedagogy delivery but not emotion capturing techniques. In a nut shell, Emotion recognition for affective learning through fusion of bio-physical and facial expression is a new strategy and couldn’t be found in any of the related research works.

Development framework-emotion detection through bio-physical and facial expression

In this section we briefly describe the core development framework used in our research. The detail explanation of Affective Computing Module shown in Figure 1 has been illustrated in details in Figure 2. Our research objective is to identify the affective state of learner using detection of emotion through two robust attributes (bio-physical and facial expression). Captured inputs have been processed and matched with learner’s six basic emotion categories. Matched emotion is accepted and processed for fusion for best result. Finally it flows to emotion evaluation section, which could customize learning pedagogy automatically. The brief of the modules are given below:

Bio-physical signal input

For Bio-Physical Input, we have used the tool ProComp5 machine which directly extracts signals from three attributes Hart Rate (HR), Skin Conductance (SC) and Blood Volume Pressure (BVP). These attributes are passed to the DTREG simulator embedded with the phase 1 of the emotion detection algorithm to obtain the function values f1, f2 and f3 which are again passed through the DTREG simulator and processed by the phase 2 of the emotion detection algorithm to generate the emotion matrix. The DTREG simulator is used as a robust simulating software tool for Predictive Modeling. Total 4608000 signals have been captured from which 18000 samples are generated. From the said sample we have group with one second and eight seconds interval and obtained data set I and data set II respectively for avoiding noisy data. A medical practitioner helped us during collection of the data.
Facial expression input

For Facial input we have used spot detection technique, automated snap shot is taken before and after the course delivery through scanning of the face involving in five different facial sections. The spot is identified before and after the course delivery as per the muscle movements and then mapped with the six defined emotions and thus identifying the learner’s emotion. Spot coordinator value \( f(x,y,\text{class}\_\text{value}) \) is passed through DTREG-SVM modeling tool for obtaining the emotion matrix.

Next, we fuse these two emotions matrices, bio-physical and facial, for accurate emotion detection. This model finds new-fangled vision for Affective Learning.

Data capturing and testing

Data capturing

For capturing facial expression, we generated a dataset consisting of 129 college students and staff ranging in age between 19 to 35 years. 45% were female, 55% were male and all were Indians. Videos were recorded in QHM495LM-3207 web camera located directly in front of the learner whose basic configurations are - lens capacity: 14 Mega pixels; output Size: 640x480; capture size 640x480; avoid flicker 50Hz; zoom 1x. Subjects have been delivered by an experimenter to perform a series of all 129 facial expressions. Image sequences from

Figure 2. Learner’s emotion sensing using bio-physical & facial input
neutral to target display were 640 by 480 pixel arrays with 8-bit precision. The only selection criterion was that a sequence be labeled as one of the 6 basic emotions (disgust, sadness, happiness, fear, anger, surprise). For collection of Bio-physical signals, the same domain has been considered concurrently where 30 sessions have conducted by maintaining all the environmental conditions, with each session lasting 20 minutes and the sampling rate 256 Hz. Finally we select 15 sessions, removing the others because of the bad placement of the bio-sensors and noise. During data collection, learners have taken number of lessons and all inputs are directly transferred to computer from ProComp5 device to generate the data file. We have completed the total process by 24 weeks. We got two sample sets for emotion detection: one sample every 1 second (data set I) and one sample every 8 seconds (data set II) for avoiding noisy data.

Data testing

A primary purpose of our data testing is to detect software or process failures to discover the defects and perform necessary corrections in the context of E-Learning platform. We have used ProComp5 and Webcam for data collection and DTREG-simulating software tool for Predictive Modeling. We have taken data for testing in two different ways concurrently, (a) By questioners and face to face emotional impact identification from learner, (b) By data collection through instruments and simulating tools. Total 18000 data were tested for Bio-physical signal inputs and 129 different data was tested for facial expression input. We had asked the questioners and noticed the emotional impact to the each learner immediately after the delivery of the session and noted out 6 basic emotions. Consciously we have tagged the input sample data with that of the corresponding learner. After processing the data through the simulating tool we matched the result from the simulator with that obtained through direct questioners and got satisfactory results. The testing result shows an approximate of accuracy of 95% accuracy. The comparison status is shown in Table 1.

<table>
<thead>
<tr>
<th>Emotions</th>
<th>Bio-physical signal – input</th>
<th>Facial expression - input</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample collected through simulator</td>
<td>Sample collected through direct questioners (face to face)</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Happiness</td>
<td>7830</td>
<td>8190</td>
</tr>
<tr>
<td>Sadness</td>
<td>5544</td>
<td>5249</td>
</tr>
<tr>
<td>Surprise</td>
<td>2502</td>
<td>2454</td>
</tr>
<tr>
<td>Anger</td>
<td>576</td>
<td>540</td>
</tr>
<tr>
<td>Disgust</td>
<td>936</td>
<td>976</td>
</tr>
<tr>
<td>Fear</td>
<td>612</td>
<td>591</td>
</tr>
<tr>
<td>Total</td>
<td>18000</td>
<td>18000</td>
</tr>
</tbody>
</table>

Algorithm, feature extraction and result analysis from bio-physical signals

In this section, we brief the methodology of capturing emotional state of the learner using bio-physical signal sensing. These features have been chosen to cover the bigger range of typically measured statistics in the bio-physical analysis for emotion database. Broadly speaking, this research indicates that a few dimensions carry most of the information about intuitive differences between emotional states (Picard, Vyzas & Healey, 2001). We developed an algorithm for emotion classification. The three parameters (f1, f2 and f3) of Bio-Physical Signals have passed through the vector to Emotion algorithm to obtain the emotion. We have identified three attributes of human body for bio-physical signal sensing, (1) Blood Volume Pressure (BVP); (2) Skin Conductance Level (SCL); and (3) Heart Rate (HR).

The affective learning model is a combination of a cognitive assessment approach to affective user modeling and a physiological approach. For the case of BVP we consider the inter-beat intervals waveform (Liu, Sourina, & Nguyen, 2010), the ProComp5 machines heart rate (H), and we have calculated inter-beat intervals, which has [1/ the inter-beat intervals (approx.)]. We applied 500-point (25 second) Hamming window (h) to form a
smoothed heart beat rate, \( b = H \ast h \), where \( n = 1, \ldots, N \), with \( N = 2000 \) (max) for Data Set I, and with \( N \) in the range of 2001 to 5000 (max) for Data Set II.

We took the means (f1):

\[
f_1 = \frac{1}{N} \sum_{n=1}^{N} b_n
\]  

(1)

The skin conductivity level signal (SCL), S contains high frequency fluctuations, which is reduced by convolving with \( h \), so \( s = S \ast h \). We have applied 500-point (25 second) Hamming window (h), from ProComp5 we have obtained \( S_n \) and \( h \). To find \( f_2 \) we put 1\text{st} and \( N\text{th} \) value as mean of the first difference of the smoothed skin conductivity. Then (f2):

\[
f_2 = \mu(S_{n+1} - S_n) = \frac{1}{N - 1} (S_N - S_1)
\]  

(2)

The respiration sensor measured expansion and contraction of the Heart Rate (HR) using a ProComp5 device. Let \( N_d \) be the number of samples collected in a session, we have collected \((N_d)= 256 \) samples in a session and \( R_n \) is respiration at \( n\text{th} \) time of sample collection in the session, R is the constant respiration / heart rate before session then we got the whole days respiration mean (f3),

\[
\mu_{R,session1} = \frac{1}{N_d} \sum_{n=1}^{N_d} R_n
\]  

(3)

Subtracted this to get \( r = R - \mu_{R,session1} \), then the respiration feature formed as

\[
f_3 = \frac{1}{N - 1} \sum_{n=1}^{N} (R_n - \mu_{R,session1})^2
\]  

(4)

In the investigation of blood-pressure changes during the session, it was found that random emotional stresses were uncovered which had potent pressure effects in normal controls, in hypertensives and in neurotics. We have taken guidance from two medical practitioners to understand the biological information. Combination of Resultant of blood volume pressure (f1), Resultant of skin conductivity level (f2) and Resultant of heart rate (f3) together can be determined the emotion of the learner. As per George Innes (Innes, Millar,&Valentine, 1959) and our observation, we propose the following range for identifying emotions in Table 2.

| Table 2. Resultant vector ranges of bio-physical signal |
|---------------------------------|-----------------|-----------------|
|                                | f1              | f2              | f3              |
| Happiness                      | 4.5 – 5.4       | 7.6 – 8.5       | 10.1 – 12       |
| Sadness                        | 1.0 – 2.4       | 4.5 – 5.5       | 0 – 2.0         |
| Surprise                       | 5.5 – 6.4       | 8.6 – 9.5       | 6.1 – 8.0       |
| Anger                          | 6.5 – 7.5       | 9.6 – 10.5      | 8.1 – 10        |
| Disgust                        | 2.5 – 3.4       | 5.6 – 6.5       | 2.1 – 4.0       |
| Fear                           | 3.5 – 4.4       | 6.6 – 7.5       | 4.1 – 6.0       |

Algorithm: Resultant vector to emotion

This algorithm is proposed to get emotion count after taking input parameter f1, f2 and f3. There is a 1D array Emotion [] of six elements. Each element of the array contents the different emotion count where, Emotion[1] = count of Happiness, Emotion[2] = count of Sadness, Emotion[3] = count of Surprise, Emotion[4] = count of Anger, Emotion[5] = count of Disgust, Emotion[6] = count of Fear. The Emotion [] array is initialized to Zero. A loop will continue until all the value of f1, f2 and f3 become NULL and within the loop array counts the emotion as per the condition applied. This algorithm returns the Emotion vector for generating confusion matrix. The complete algorithm is given below:

Input: f1, f2 & f3
Output: Emotion [6]
Emotion [] = {0, 0, 0, 0, 0, 0};
do{
if (f1 between(4.5,5.4) && f2 between(7.6,8.5) && f3 between(10.1,12))
    Emotion [1]++; // Happiness count
if (f1 between(1.0,2.4) && f2 between(4.5,5.5) && f3 between(0,2.0))
    Emotion [2]++; // Sadness Count
if (f1 between(5.5,6.4) && f2 between(8.6,9.5) && f3 between(6.1,8.0))
    Emotion [3]++; // Surprise count
if (f1 between(6.5,7.5) && f2 between(9.6,6.5) && f3 between(8.1,10))
if (f1 between(2.5,3.4) && f2 between(5.6,6.5) && f3 between(2.1,4.0))
if (f1 between(3.5,4.4) && f2 between(6.6,7.5) && f3 between(4.1,6.0))
    Emission [6]++; // Fear count
}while (f1&&f2&&f3!= NULL);
return (Emotion);

We have used DTREG (https://www.dtreg.com), robust simulating software for Predictive Modeling and Forecasting, which extracts useful information from a set of data values. SVM techniques have been used for developing predictive modeling, and there is an art of selecting and applying the best method for a particular situation. We obtain our data range from Table 1 for obtaining the result of emotion array. The functions f1, f2, f3 and “Resultant vector to Emotion” functions are supplied to the DTREG simulator. The final predictive value of emotion matrix (Happiness, Sadness, Surprise, Anger, Disgust, and Fear) is obtained in two phases through DTREG modeling simulator. In the first phase, we obtained f1, f2 and f3 from the data file (data I and data II). In the second phase we obtained the emotion classification from f1, f2 and f3 as per our proposed algorithm and the total process is shown in Figure 3. We have taken 18,000 samples and identified the class of emotion count report summary as shown in Table 2.

For emotion recognition, we have taken 30 sessions, with each session lasting 20 minutes and the sampling rate of 256 Hz. Finally, we select 15 sessions, removing the others due to the bad placement of the bio-sensors and noise. We obtained a total of 4608000 samples (15 session * 20minutes/session * 60 seconds/minutes *256 samples/seconds = 4608000).

Table 3. Brainwaves and their relationship with emotion as per the simulator documentation

<table>
<thead>
<tr>
<th>Wave type</th>
<th>Frequency</th>
<th>When wave is dominant</th>
</tr>
</thead>
<tbody>
<tr>
<td>δ Delta</td>
<td>0-4 Hz</td>
<td>Deep sleep</td>
</tr>
<tr>
<td>θ Theta</td>
<td>5-8 Hz</td>
<td>Creativity, dream sleep drifting thoughts</td>
</tr>
<tr>
<td>α Alpha</td>
<td>9-13 Hz</td>
<td>Relaxation, calmness, abstract thinking</td>
</tr>
<tr>
<td>Low β Beta</td>
<td>14-20 Hz</td>
<td>Relaxed focus. High alertness, mental activity. Agitation anxiety.</td>
</tr>
<tr>
<td>High β Beta</td>
<td>21-40 Hz</td>
<td></td>
</tr>
</tbody>
</table>

Such a big data set makes training and classification very time-consuming. To make it more efficient, and as emotion won’t change so frequently, we merged n samples into 1(one) sample. When n = 256, then we have 1 sample every 1 second; when n = 2048, then we have 1 sample every 8 seconds.

The data set I and data set II has been forwarded to DTREG-SVM modeling simulator for the detection of emotion followed by SVM based classification. The parameters of the machine are given in Table 3 and 4 and the corresponding data set in Table 5 and 6.
Table 4. Parameters assigned for Data Set I and II

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training_file</td>
<td>train.matrix.txt</td>
<td>zero_mean</td>
<td>false</td>
</tr>
<tr>
<td>matrix_from_file</td>
<td>False</td>
<td>normalize</td>
<td>true</td>
</tr>
<tr>
<td>variance_one</td>
<td>False</td>
<td>coefficient</td>
<td>1</td>
</tr>
<tr>
<td>Constant</td>
<td>10</td>
<td>bias</td>
<td>0</td>
</tr>
<tr>
<td>Power</td>
<td>1</td>
<td>width_factor</td>
<td>1</td>
</tr>
<tr>
<td>Radial</td>
<td>False</td>
<td>add_diag</td>
<td>0</td>
</tr>
<tr>
<td>two_squared_width</td>
<td>0</td>
<td>thresh_type</td>
<td>percent</td>
</tr>
<tr>
<td>feature_select</td>
<td>None</td>
<td>sum_of_weights</td>
<td>0.00742884</td>
</tr>
<tr>
<td>Fthreshold</td>
<td>0</td>
<td>negative_constraint</td>
<td>0</td>
</tr>
<tr>
<td>positive_constraint</td>
<td>0</td>
<td>positive_diagonal</td>
<td>0.66</td>
</tr>
<tr>
<td>constrain_weights</td>
<td>False</td>
<td>convergence_threshold</td>
<td>1e-06</td>
</tr>
<tr>
<td>negative_diagonal</td>
<td>0.44</td>
<td>objective</td>
<td>0.236593</td>
</tr>
<tr>
<td>Seed</td>
<td>1342932938</td>
<td>iterations</td>
<td>256</td>
</tr>
</tbody>
</table>

Note. The values have been assigned as per the standard of the ProComp5 device.

The data is received from ProComp5, which is passed to the simulator for obtaining f1, f2 and f3. Data Set I is obtained for intervals of 1 second. First 12 sample values are shown in Table 5.

Table 5. Data Set I (first 12 values)

<table>
<thead>
<tr>
<th>Sample no.</th>
<th>Class</th>
<th>Weight</th>
<th>train_classification</th>
<th>train_discriminant</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.0000</td>
<td>1</td>
<td>1.3870</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.0000</td>
<td>1</td>
<td>1.1490</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.1211</td>
<td>1</td>
<td>0.9200</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0.2556</td>
<td>1</td>
<td>0.8289</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.3847</td>
<td>1</td>
<td>0.7433</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0.4238</td>
<td>1</td>
<td>0.7140</td>
</tr>
<tr>
<td>7</td>
<td>-1</td>
<td>-0.7000</td>
<td>-1</td>
<td>-0.6918</td>
</tr>
<tr>
<td>8</td>
<td>-1</td>
<td>-0.2233</td>
<td>-1</td>
<td>-0.9029</td>
</tr>
<tr>
<td>9</td>
<td>-1</td>
<td>-0.2016</td>
<td>-1</td>
<td>-0.9166</td>
</tr>
<tr>
<td>10</td>
<td>-1</td>
<td>-0.0530</td>
<td>-1</td>
<td>-0.9804</td>
</tr>
<tr>
<td>11</td>
<td>-1</td>
<td>-0.0561</td>
<td>-1</td>
<td>-0.9811</td>
</tr>
<tr>
<td>12</td>
<td>-1</td>
<td>-0.0511</td>
<td>-1</td>
<td>-0.9187</td>
</tr>
</tbody>
</table>

Data Set II is obtained based on a periodic interval of 8 seconds to avoid noisy data. First 12 sample values are shown in Table 6.

Table 6. Data Set II (first 12 values)

<table>
<thead>
<tr>
<th>Sample no.</th>
<th>Class</th>
<th>Weight</th>
<th>train_classification</th>
<th>train_discriminant</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>-0.0032</td>
<td>1</td>
<td>1.3636</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.0032</td>
<td>1</td>
<td>1.1724</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.1179</td>
<td>1</td>
<td>0.8966</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0.2588</td>
<td>1</td>
<td>0.8523</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.3815</td>
<td>1</td>
<td>0.7199</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0.4270</td>
<td>1</td>
<td>0.7374</td>
</tr>
<tr>
<td>7</td>
<td>-1</td>
<td>-0.7032</td>
<td>-1</td>
<td>-0.7152</td>
</tr>
<tr>
<td>8</td>
<td>-1</td>
<td>-0.2201</td>
<td>-1</td>
<td>-0.8795</td>
</tr>
<tr>
<td>9</td>
<td>-1</td>
<td>-0.2048</td>
<td>-1</td>
<td>-0.9400</td>
</tr>
<tr>
<td>10</td>
<td>-1</td>
<td>-0.0498</td>
<td>-1</td>
<td>-0.9570</td>
</tr>
<tr>
<td>11</td>
<td>-1</td>
<td>-0.0593</td>
<td>-1</td>
<td>-1.0045</td>
</tr>
<tr>
<td>12</td>
<td>-1</td>
<td>-0.0479</td>
<td>-1</td>
<td>-0.8953</td>
</tr>
</tbody>
</table>
We use the values of f1, f2 and f3 and obtain the results, which are grouped in two classes for learning style identification and are given in Table 8.

Table 8. Emotion - Group - Decision taken

<table>
<thead>
<tr>
<th>Emotion detected</th>
<th>Grouping</th>
<th>Learning style detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happiness, surprise</td>
<td>Positive Group</td>
<td>Lesson Understood</td>
</tr>
<tr>
<td>Sadness, fear, disgust, anger</td>
<td>Negative Group</td>
<td>Lesson Not Understood</td>
</tr>
</tbody>
</table>

The algorithm for Resultant vector to Emotion conversion generates the distribution of emotion from all the samples. The results obtained after classifications are shown in the Table 9.

Table 9. The emotion distribution for all 18000 samples through biophysical signals

<table>
<thead>
<tr>
<th>Emotions</th>
<th>Number of sample</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happiness</td>
<td>7830</td>
<td>44%</td>
</tr>
<tr>
<td>Sadness</td>
<td>5544</td>
<td>31%</td>
</tr>
<tr>
<td>Surprise</td>
<td>2502</td>
<td>14%</td>
</tr>
<tr>
<td>Anger</td>
<td>576</td>
<td>3%</td>
</tr>
<tr>
<td>Disgust</td>
<td>936</td>
<td>5%</td>
</tr>
<tr>
<td>Fear</td>
<td>612</td>
<td>3%</td>
</tr>
<tr>
<td>Total</td>
<td>18000</td>
<td>100%</td>
</tr>
</tbody>
</table>

The confusion matrix contains information about actual and predicted classifications done by a classification system. As per the above prediction technique a confusion matrix has been generated for Bio-physical classification. Each cell gives the number of cases of the row’s emotion classified as the column’s emotion. Rightmost column gives the overall mean accuracy for each row emotion. This matrix provides a quantitative performance representation for each classifier in terms of class recognition. The main diagonal elements indicate the total number of samples in the emotion class correctly recognized by the system. Table 10 shows the performance of the emotion recognition system based on the biophysical expressions when all the samples are evaluated. The overall performance of this classifier is 68.98 %. Happiness and Anger are recognized with very high accuracy (84.12 and 81.13 % respectively).

Table 10. Confusion matrix of the emotion recognition system based on biophysical signal

<table>
<thead>
<tr>
<th>Predicated</th>
<th>Happiness</th>
<th>Surprise</th>
<th>Disgust</th>
<th>Anger</th>
<th>Fear</th>
<th>Sadness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happiness</td>
<td>84.12</td>
<td>4.63</td>
<td>4.49</td>
<td>0</td>
<td>2.81</td>
<td>3.95</td>
</tr>
<tr>
<td>Surprise</td>
<td>10.56</td>
<td>63.6</td>
<td>0</td>
<td>13.85</td>
<td>0</td>
<td>11.99</td>
</tr>
<tr>
<td>Disgust</td>
<td>4.13</td>
<td>0</td>
<td>63.72</td>
<td>8.45</td>
<td>0</td>
<td>23.7</td>
</tr>
<tr>
<td>Anger</td>
<td>0</td>
<td>5.33</td>
<td>6.67</td>
<td>81.13</td>
<td>3.33</td>
<td>3.54</td>
</tr>
<tr>
<td>Fear</td>
<td>0</td>
<td>0</td>
<td>7.35</td>
<td>17.66</td>
<td>64.66</td>
<td>10.33</td>
</tr>
<tr>
<td>Sadness</td>
<td>0</td>
<td>19.67</td>
<td>0</td>
<td>13.34</td>
<td>10.33</td>
<td>56.66</td>
</tr>
</tbody>
</table>

Note. The accuracy (AC) is the proportion of the total number of predictions. Overall Performance = (84.12+63.6+63.72+81.13+64.66+56.66)*100/600 = 68.98 %.
Algorithm and feature extraction and result analysis from facial expression

Facial expressions give important and acute clues about emotions. Therefore, the features used are typically based on local spatial position or displacement of specific points and regions of the face. Our objective is to detect affective state of learner through facial emotion (Raouziou, Ioannou, Karpouzis, Tsapatsoulis, Kollias, & Cowie, 2003); hence we choose the spot detection technique. Here we brief the complete process of facial expression extraction. The detection technique is performed in the client side, which minimizes the server side workload. We don’t detect the full face; only spot movement detection is enough for emotion detection. Emotion can easily be detected from the same or different group of learners because of the facial pattern. In the system based on visual information, the spatial data is collected from markers in each of the frames of the video. A 6-dimensional feature vector per sentence is used as input to the classifier. The facial expression system, which is shown in Figure 4, is described below. After capturing the motion data, it has been normalized.

Figure 4. Five areas of the face has been considered in this study (random sampling) and facial points before and after course delivery.

We have extracted the set of values from the facial spots as shown in Figure 4(c). Identified the spots throughout the face and as per the muscle movements and it has been divided into the 5 blocks. Before and after the delivery the spots have been taken and are mapped for detection of pattern using SVM. The black points are considered in class (+1), which were taken before the session, and the blue points are considered in class (-1), which were taken after the session. We have taken training data, which is a set of n points.

Positive: \( \langle w \cdot x \rangle + b = +1 \)
Negative: \( \langle w \cdot x \rangle + b = -1 \)
Hyper plane: \( \langle w \cdot x \rangle + b = 0 \)

We have to find the unknowns, w and b by expending the equations:

\[
\begin{align*}
  w_1x_1 + w_2x_2 + b &= +1 \quad \text{(5)} \\
  w_1x_1 + w_2x_2 + b &= -1 \quad \text{(6)} \\
  w_1x_1 + w_2x_2 + b &= 0 \quad \text{(7)}
\end{align*}
\]

Data segments of 1200 seconds in length were taken from each of the signals. Happiness and Sadness are the main opposite Learner’s mode which has the major difference in frequency. We have generated the training data for facial emotions before and after course delivery. One dataset contained 129 values of which first ten rows of the training data is shown in Table 11. For all training snaps same set of tables have been generated.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Class</th>
<th>X</th>
<th>Y</th>
<th>Class</th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>SL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>+1</td>
<td>15</td>
<td>71</td>
<td>-1</td>
<td>17</td>
<td>65</td>
</tr>
<tr>
<td>2</td>
<td>+1</td>
<td>20</td>
<td>61</td>
<td>-1</td>
<td>23</td>
<td>58</td>
</tr>
<tr>
<td>3</td>
<td>+1</td>
<td>25</td>
<td>51</td>
<td>-1</td>
<td>30</td>
<td>57</td>
</tr>
<tr>
<td>4</td>
<td>+1</td>
<td>34</td>
<td>42</td>
<td>-1</td>
<td>39</td>
<td>44</td>
</tr>
<tr>
<td>5</td>
<td>+1</td>
<td>47</td>
<td>38</td>
<td>-1</td>
<td>54</td>
<td>41</td>
</tr>
<tr>
<td>6</td>
<td>+1</td>
<td>61</td>
<td>37</td>
<td>-1</td>
<td>68</td>
<td>39</td>
</tr>
<tr>
<td>7</td>
<td>+1</td>
<td>75</td>
<td>38</td>
<td>-1</td>
<td>85</td>
<td>41</td>
</tr>
<tr>
<td>8</td>
<td>+1</td>
<td>91</td>
<td>46</td>
<td>-1</td>
<td>100</td>
<td>47</td>
</tr>
<tr>
<td>9</td>
<td>+1</td>
<td>100</td>
<td>52</td>
<td>-1</td>
<td>111</td>
<td>54</td>
</tr>
<tr>
<td>10</td>
<td>+1</td>
<td>108</td>
<td>61</td>
<td>-1</td>
<td>47</td>
<td>58</td>
</tr>
</tbody>
</table>

We have used the non linear SVM algorithm to construct a hyperplane as the decision plane, which separates the positive (+1) and negative (-1) classes with the largest margin. It is related to the minimization of the VC
dimension (VapnikChervonenkis Dimension) of SVM, which measures the capacity of a hypothesis space. The block diagram of generation of emotion matrix from facial expression is shown in Figure 5.

![Figure 5. Generation of emotion matrix from facial spot](image)

Table 12. The emotion distribution for all 129 samples through facial signals

<table>
<thead>
<tr>
<th>Emotions</th>
<th>Number of sample</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happiness</td>
<td>52</td>
<td>40%</td>
</tr>
<tr>
<td>Sadness</td>
<td>40</td>
<td>31%</td>
</tr>
<tr>
<td>Surprise</td>
<td>15</td>
<td>12%</td>
</tr>
<tr>
<td>Anger</td>
<td>11</td>
<td>9%</td>
</tr>
<tr>
<td>Disgust</td>
<td>9</td>
<td>7%</td>
</tr>
<tr>
<td>Fear</td>
<td>2</td>
<td>2%</td>
</tr>
<tr>
<td>Total</td>
<td>129</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 13. Confusion matrix of the emotion recognition system based on facial emotion

<table>
<thead>
<tr>
<th></th>
<th>Happiness</th>
<th>Surprise</th>
<th>Disgust</th>
<th>Anger</th>
<th>Fear</th>
<th>Sadness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happiness</td>
<td>80.12</td>
<td>4.63</td>
<td>8.49</td>
<td>0</td>
<td>0</td>
<td>6.76</td>
</tr>
<tr>
<td>Surprise</td>
<td>12.56</td>
<td>66.67</td>
<td>0</td>
<td>13.58</td>
<td>7.19</td>
<td>0</td>
</tr>
<tr>
<td>Disgust</td>
<td>0</td>
<td>0</td>
<td>65.72</td>
<td>10.45</td>
<td>0</td>
<td>23.83</td>
</tr>
<tr>
<td>Anger</td>
<td>0</td>
<td>3.33</td>
<td>6.67</td>
<td>80.1</td>
<td>3.33</td>
<td>6.57</td>
</tr>
<tr>
<td>Fear</td>
<td>0</td>
<td>0</td>
<td>6.66</td>
<td>15.66</td>
<td>65.33</td>
<td>12.35</td>
</tr>
<tr>
<td>Sadness</td>
<td>0</td>
<td>6.67</td>
<td>23.66</td>
<td>0</td>
<td>13</td>
<td>56.67</td>
</tr>
</tbody>
</table>

*Note. Overall performance = (80.12+66.67+65.72+80.1+65.33+56.67) * 100 / 600 = 69.10%.*

We have taken the spot coordinates as the input for the two different classes. One class of (+1) which are taken before the learning session and another class of (-1) which are taken after the learning session and 129 pairs (black and blue) are taken. DTREG-SVM simulator applies the non-linear SVM algorithm to obtain the hyperplane to identify the pattern and recognize emotion (Duda, Hart, & Stork, 2000). The total classified data has been bundled in the following Table 12.

Table 13 shows the performance of the emotion recognition system based on facial emotion when all the samples are used. The overall performance of this classifier was 69.10%. Anger and Happiness are recognized with very high accuracy (80.12 and 80.1 % respectively).

**Fusion of Bio-physical and facial emotions and automatic selection of lesson**

Decision level fusion, in the context of a verification system, has been posed as a binary hypothesis-testing problem involving multiple classifiers of emotions. In decision level fusion, each classifier operating under a binary hypothesis, pertained a threshold on the match score and renders its decision regarding the presence (= 1) or absence (= 0) of an authentic emotion. The decisions from multiple classifiers are then fused in order to generate the final decision of emotion. The block structure of fusion is given below in Figure 6.

![Figure 6. Decision-Level Fusion framework](image)
In our research, we obtained two confusion matrixes, which we got from bio-physical emotion and facial emotion and fused these two matrixes for better classification. Achieving optimality at the decision level, however, involves the selection of optimal decision thresholds and a fusion rule that minimize the classification error. Achieving optimality at the decision level, however, involves the selection of optimal decision thresholds and a fusion rule that minimize the classification error (Duda, Hart, & Stork, 2001; Pantic, & Rothkrantz, 2003). Two confusion matrices of bio-physical and facial emotion are fused and to obtain the fusion matrix as shown in Table 14.

E-learning environments can take advantage of the different forms of learning by recognizing the pedagogy of each individual student using the system and adapting the content of courses to match this style. The method is based on artificial neural networks (ANNs) (Keefe, 1979). In the proposed approach, neural networks are used to recognize learners’ learning styles based upon the actions they have performed in an E-Learning system. As per the detection of affective state, the system will suggest the lesson to the individual learner. The system can select the learning pedagogy of the learner and neuro-fuzzy logic is required to implement such methodology. The flow diagram is shown in Figure 7.

**Table 14.** Fusion matrix of the facial emotion and biophysical signal

<table>
<thead>
<tr>
<th>Actual Emotions</th>
<th>Predicted Emotions</th>
<th>Happiness</th>
<th>Surprise</th>
<th>Disgust</th>
<th>Anger</th>
<th>Fear</th>
<th>Sadness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happiness</td>
<td>89.22</td>
<td>4.63</td>
<td>2.49</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3.66</td>
</tr>
<tr>
<td>Surprise</td>
<td>10.56</td>
<td>64.6</td>
<td>0</td>
<td>12.23</td>
<td>0</td>
<td>0</td>
<td>12.61</td>
</tr>
<tr>
<td>Disgust</td>
<td>4.66</td>
<td>0</td>
<td>63.72</td>
<td>11.45</td>
<td>0</td>
<td>0</td>
<td>20.17</td>
</tr>
<tr>
<td>Anger</td>
<td>0</td>
<td>5.33</td>
<td>0</td>
<td>82.11</td>
<td>3.33</td>
<td>0</td>
<td>9.23</td>
</tr>
<tr>
<td>Fear</td>
<td>0</td>
<td>0</td>
<td>7.7</td>
<td>17.98</td>
<td>64.66</td>
<td>9.66</td>
<td></td>
</tr>
<tr>
<td>Sadness</td>
<td>0</td>
<td>9.67</td>
<td>0</td>
<td>13.34</td>
<td>10.33</td>
<td>66.66</td>
<td></td>
</tr>
</tbody>
</table>

*Note. Overall Performance = (89.22+64.6+63.72+82.11+64.66+66.66)*100/600 = 71.83%.

**Conclusion**

Using facial and biophysical signals, this study explores the emotion evolution during learning, and proposes an Affective E-Learning model. We have developed learner’s emotion sensing methodology using bio-physical and facial expression with efficient classification. Facial and biophysical data are gathered from subjects over twenty four weeks in a natural and close-to-real world setting. The best-case classification rate (89.22%) yielded
through decision-level fusion of facial and bio-physical data, opening up the possibilities for instructors to understand the emotional states of remote learners. The implication of our research is that, it needs to be implemented and commercially shaped through real time Learning Management System (LMS) to make it Affective LMS. We also wish to work with other signals like body gesture, speech etc. to increase the maximum competency in human computer interaction for building up an Affective E-Learning System.

Acknowledgements

We are thankful to all staffs the Lab. We especially thankful to Dr. D Malhotra, and Dr. M Sultan helped us for handling and retrieving data from ProComp5 device.

References


Exploring Teaching Programming Online through Web Conferencing System: The Lens of Activity Theory

Ünal Çakıroğlu1, Mehmet Kokoc2, Elvan Kol3 and Ebru Turan4

1Department of Computer Education and Instructional Technologies, Karadeniz Technical University, Turkey // 2Center for Research and Application in Distance Education, Karadeniz Technical University, Turkey // 3Ministry of Education, Turkey // 4Department of Computer Education and Instructional Technologies, Giresun University, Turkey // cakiroglu@ktu.edu.tr // mkokoc@ktu.edu.tr // elvankol61@hotmail.com // ebru.turan.guntepe@giresun.edu.tr

*Corresponding author

(Submitted June 16, 2015; Revised November 11, 2015; Accepted January 8, 2016)

ABSTRACT

The purpose of this qualitative study was to understand activities and behaviors of learners and instructor in an online programming course. Adobe Connect web conferencing system was used as a delivery platform. A total of fifty-six sophomore students attending a computer education and instructional technology program (online) participated in this study. Data were collected through observations of twenty-two online course recordings and semi-structured interviews. The findings were discussed through Mwanza’s Activity notations based on Activity Theory. The results indicated that, by using web conferencing tools, students could develop programming knowledge through the learning tasks by interacting each other. The instructor acted as a key role in constructing the online learning community. Opportunities of getting instant feedback, acting in online collaborative activities and sharing features of the system contributed to students’ learning. Also, it is suggested that; both conceptual and strategic structures of the programming courses should be taken into consideration while using web conferencing.

Keywords

Teaching programming online, Activity theory, Online learning, Web conferencing system

Introduction

Researchers indicate that instructors face several problems during the programming courses (Yang et al., 2015). Abstract nature of the programming task (Kátai & Toth, 2010), students’ readiness for problem-solving (Ozdinc & Altun, 2014) and, repetitive failures in the instructional process (Law, Lee & Yu, 2010) are the frequently faced problems. In recent years, the increasing demand for programming courses has led many institutions to utilize online environments for delivering these courses. The attributes of the technologies and instructional methods have crucial roles in the effectiveness the courses. In this sense, Levy (2006) suggested that designing and implementation of the activities in instructional process required considerable attention to organize all components that affect the learning outcomes. Therefore, exploring learning contexts in online programming course may guide online instructors.

The analysis of relationships between the components in learning contexts may facilitate understanding how online programming teaching learning process occurs. Thus, there is a need to explore the process in a philosophical framework. In this regard, Activity Theory (AT) is suggested as a framework to clarify the activities within a course after a certain amount of time (Jonassen & Rohrer-Murphy, 1999; Karakus, 2013). In this study, AT was used for the purpose of exploring the context of teaching programming online considering technology, people, and learning environment.

Activity theory

AT offers a framework for modeling human behaviors in the activities (Kuutti, 1996). Human interactions with other components are recognized using activity as basic analysis unit. AT takes its origins from both Leontev’s notion of activity and Vygotsky’s idea about the tool mediation. This idea is generally considered as the mediational model of human interactions with environment as seen in Figure 1 (Mwanza, 2001).

The subject and object components have a mediated relationship through the use of tools each other. Leontev (1978) developed a hierarchical model for human activity by addressing social and cultural mediation. Based on Vygotsky’s idea, Engeström (1987) expanded the relationship between subject and object in terms of social and...
cultural aspects of human activity. He developed the activity triangle model, which presents interactions within the activities in a more comprehensive manner, by adding components such as rules, division of labour and tools.

The activity triangle model incorporates subjects, object, and community components as well as mediators of human activity, tools, rules, and the division of labour. These components are briefly defined in Table 1.

<table>
<thead>
<tr>
<th>AT Components</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>The individual or a group engaging to any activity generally considered as the subject (Jonassen &amp; Rohrer-Murphy, 1999).</td>
</tr>
<tr>
<td>Object</td>
<td>Object is motivation of the action and goal of the activity (Engeström, 2001).</td>
</tr>
<tr>
<td>Outcome</td>
<td>Outcome is a concrete or abstract artifact of the activity (Karakus, 2013). Objects can be transformed into outcomes.</td>
</tr>
<tr>
<td>Tool</td>
<td>A tool can be anything used in the transformation of objects into outcomes (Kuutti, 1996).</td>
</tr>
<tr>
<td>Community</td>
<td>Community refers to the context and people with whom the subject interacted (Karakus, 2013).</td>
</tr>
<tr>
<td>Rules</td>
<td>Rules can be formal, informal and technical norms or conventions of community (Barab, Schatz, &amp; Scheckler, 2004).</td>
</tr>
<tr>
<td>Division of Labour</td>
<td>Division of Labour refers to responsibilities of community members (Karakus, 2013). It is generally considered due to the roles which the subjects act in the community (Baran &amp; Cagiltay, 2010).</td>
</tr>
</tbody>
</table>

Within these definitions, Kuutti (1996), who analyzed Engeström’s AT triangle, pointed out that the relationships between subject and object is mediated through tools; subject and community is mediated through rules; and object and community is mediated through division of labour. The AT components are in a continuous interaction during activity process and create the contexts including deliberate actions that they develop to achieve the objectives (Sam, 2012).

**Activity theory as an analysis framework**

Social learning perspective indicates that learning occurs as a result of interactions with the human and objects (Driscoll, 2005; Vygotsky, 1978). According to Vygotsky, social interaction and cultural context have an effect on cognition and learning. Likewise, AT suggests that learning emerges from human activity in a complex activity with interaction among subjects, objects, and tools (Engeström, 1987; Jonassen & Rohrer-Murphy, 1999). For understanding and analyzing complex human learning activities in a context, AT may be an analytical framework for analyzing the system (Yamagata-Lynch, 2013).

Activities consist of actions that are chains of operations (Jonassen & Rohrer-Murphy, 1999) and subject performs to transform object into outcomes in an activity system (Karakus, 2013, p. 152). Bodker (1997) described these actions and operations that are mediated by tools, words, and/or cultural signs. In this sense, Hasan and Crawford (2003) adapted Generic Human Activity from AT which includes the components as environment, tools, people, actions, operations, and outcome. In the framework, activities are revisited for online learning so that it comprises such components as online learning system, tools, learners and teachers, actions-operations, and learning outcomes. Baran and Cagiltay (2010) used sub-components that they call four sub-activity systems in their study (subject-tool-object, subject-rules-object, subject-community-object, and subject-division of labour-object) in which subject and object remained stable since they are the main elements of the activity system while the mediators (tools, rules, community, and division of labour) varied.
As different from the others, Mwanza (2001) followed an eight-step model, and suggested to use AT components effectively. The model suggests, first; identify and clearly show what sort of activity will be adopted. Then, specify why this activity will take part depending on the purpose or the result of the activity. Thereafter, define the subjects in the activity and make a decision about the responsibilities of the tasks. It is also necessary to decide which tools will be utilized and which rules will be set. Finally, the labour needs to be shared among the community in the activity, and the expectations from this activity should be clearly stated. Mwanza (2001) in the Activity Notation splits the activity triangle into six smaller and manageable units (triangles) to make AT more understandable which is shown in Figure 2.

Different forms of AT demonstrate its expandability to explore the various dynamics of learning environments. In online learning literature, it seems that there is an agreement that different forms of AT may be used when describing learning activities performed to achieve a desired learning outcome (Peña-Ayala et al., 2014; Sancho-Thomas et al., 2009). Thus, in this study, it is hypothesized that AT framework may be used to explore an online synchronously delivered programming course.

Research problem

The nature of the course, instructor and learner characteristics, tools, activities in communication, and collaboration were mostly taken into consideration in online learning research. In this study, online lessons were delivered via web conferencing system. The learning tasks were considered as basic activities of the learning community. So, the study aims to reveal the instructor’s and learners’ experiences in the activities in order to evaluate learning outcomes by providing Mwanza’s (2001) conceptualization of AT as a framework. The following sub-questions were guided the research study within context of the online programming course (learners, instructor, learning environment, etc.).
- How did the instructional activities take place within interactions among learners, instructor and the web conferencing system?
- What were the learning outcomes of a programming course delivered via web conferencing according to the experiences and perspectives of the learners?

Method

This qualitative study provides contextual descriptions to understand human behaviors in online learning activities through web conferencing system. In this regard, the descriptive potential of qualitative studies for context was utilized. In the same vein, Jonassen and Roher-Murphy (1999) recommended that qualitative research and AT work together because activities can only be meaningfully understood in context. AT also employs a variety of qualitative methods such as interviews, observations, and document analysis for different perspectives to understand the activities (Roth & Lee, 2007).
**Participants**

A total of 56 undergraduate students (35 male, 21 female) enrolled to the Computer Education and Instructional Technology (CEIT) program at a major university participated in the study. The students were candidate teachers between 18 and 24 years old who had limited prior knowledge about programming. The participants were selected with purposeful sampling method.

**Course context**

In the program, most of the courses were delivered online, through web conferencing system (Adobe Connect). The study took place in Introductory Programming Languages which was a mandatory course and delivered via Adobe Connect as well. The goal of the course was to provide students understanding to solve problems by using programming languages including basic concepts, data types, control structures, logical operators, basic functions and loops. Active participation, endeavoring with exploratory tasks as homework, and efficient group work were implemented to support developing programming knowledge.

In this study; when instructor was handling mainly conceptual and structural topics he used various teaching methods. For example, while he was teaching conditions or operation topics, he referred several cases. He provided discussions on examples related to daily life problems. Students shared their ideas as voice and video as well as text with permission and involvement in the chat window. While teaching basic controls and loops, instructor presented samples and asked students to suggest alternative solutions. Students were allowed to use hand tool and shared their suggestions simultaneously. While discussing alternative solutions instructor provided instant feedback to help students reaching a common decision. Then simultaneously students, themselves wrote code pieces and instructor asked one or more students to share the codes on their desktop. This helped students to compare their own codes with the shared ones. In some cases, students set cooperative groups for developing efficient codes. After finding an efficient result, they shared it with others as a decision of the group. Particularly on the homework, which requires strategic knowledge, instructor selected students in participant list and asked them to share their ideas in written, audio and video format. Specifically, when the instructor found some ideas interesting or valuable, he gave live video permission to students to provide their ideas with using webcams. Instructor sometimes directed students to discuss around the valuable ideas until to reach a negotiation. By this way, he motivated students and improved interaction in order to provide feelings that they are in a real classroom. Consequently, the teaching methods and the tools used by the participants were varied depending on the nature of the topics. It is seen that; by using Adobe Connect as a delivery tool, the activities in the lessons may include various interactions between elements (students, instructor, system, teaching-learning process, course objectives) of AT at a certain level. So, we can have a systematic view in this learning community by using AT. In this regard, the affordances of web conferencing system were presented below.

**Web conferencing system**

Adobe Connect, which is one of the web conferencing systems, provides synchronous interaction in an online environment.

![Figure 3. The interface of the web conferencing system](image-url)
The system includes tools for audio-visual communication and interaction between student and instructor and student and student (See Figure 3). Users can share a variety of content types including files, images, presentations, audio, video and their desktop view including applications synchronously. Students can send and receive public and private messages via a chat tool. They can also share their voice and video. In this environment, about 10 to 12 students can participate in group discussions by connecting the system concurrently with voice and video.

Data collection procedure

In the contextual descriptions in which AT is adopted as a framework, most of the studies (Baran & Cagiltay, 2010; Heo & Lee, 2013; Park et al., 2013; Thomas & McRobbie, 2013) have a tendency to use qualitative data collection instruments regardless of the structure and the number of the participants. Thus, we used observations and semi-structured interviews to collect data. In this study, 22 online course records (90-110 min.) were observed. In order to analyze activities in the videos, an observation form including various items based on Mwanza’s model was developed by the researchers.

In addition, the semi-structured interview questions were developed through the opinions of two faculty members. One is experienced in Adobe Connect for years; other has experiences on distance education organization. The interviews were conducted with five participants (two male and three female) to reveal students’ behaviors, perspectives, and interactions in the online programming learning environment. The participants were selected because they had no absence throughout the academic year, participated in all of the exams, and they were volunteered to participate in the interviews.

Interview questions were developed considering the context of the study and observation results. The interview protocol consisted of six open-ended questions inquiring how students perceived possible effects of instructional activities using system tools and, what the students and instructors’ experiences were in online programming course. The interviews were conducted in a quite study room in university with an average length of 18 minutes and they were recorded via voice recorder with the informed consent of the participants. The interviews transcribed verbatim for data analysis.

Data analysis

Mwanza’s Activity Notation in this study was used to explain the interactions between subject and community (Mwanza, 2002). There are three essential rules of the notation: (1) “actor” represented by the subject or community components, (2) “mediator” represented by the tools, rules, or division of labour components of the triangle model, and (3) “object” which is focused on in the activity. It is obvious that the two fundamental components, subject and community, are provided through the mediator components (tools, rules, division of labour) of the relationship with the object, considering that object is the component which is focused in the activity. In this regard, it can be stated that subject, mediator, and object combinations can be used for exploring the context in this study. These combinations are shown in Table 2.

<table>
<thead>
<tr>
<th>Table 2. Mwanza’s Activity Notation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Subject</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Community | Tools | Object |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Rules</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Division of Labour</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Regarding the combinations, observations of each item was rated as rarely (R), occasionally (O), and frequently (F) relying on the nature of the course. These ratings indicate the frequency of the observed actions. Thereby, frequency of an action varies depending on the context. For example, desktop share including opening windows and applications and getting feedback are faced less often in an online programming, so, these actions are assessed as “frequently.” In the addressing process, the opinions of the two instructors, who are experts in online programming instruction were considered.
For the interviews, the questions were prepared beforehand, depending on the answers during the interviews or additional questions were asked to clarify the interviewees’ answers. In addition, sufficient attention was paid to ask a single question at once, prepare alternative questions that reflect the interview questions in different ways. The questions were asked in a conversational tone so that students could feel comfortable in the interview. The interview records were transcribed by researchers and interpreted within the AT sub-components. The researchers coded the statements individually and then they discussed with each other until they come to exact agreement on each codes.

Results

The results for analysis of video records were illustrated in Figure 4 in the form of codes, including the relationships of the dimensions as themes with taking “object” as the main goal.

![Figure 4. Frequency of AT sub-components in the recorded courses](image)

In order to analyze the activities within AT framework, activity types and the people engage in activity, rules for the activity, the reasons for action, the outcome from the activity, and the community in which the activity occurs should be presented. So, considering web conferencing as an activity system in which learning occurs, results from the 22 videos were presented through sub-components within the framework in Figure 5.

In this study subject in the web conferencing system is the students (candidate teachers) which are the driving characters in the activities. Students were individual or group of actors engaged in the activities. They have received basic ICT course including some practical knowledge about using basic Office components in the previous year. Previous course included only limited concepts about programming components such as variable, memory, data types etc. The object is current presentation and sharing of the information related to the learning objectives. Learning outcomes are resulting artifacts of the programming activities, in other words transformation of object is outcome. Three steps of programming (analyze, design and code) were desired outcomes of the study as in many other introductory programming courses. To produce desired outcome, students change the present knowledge to further knowledge via connections or interactions with subjects. In the study, learning programming emerges from the interactions of students and instructors in the activities within a context in Adobe Connect system. They use web conferencing tools to transform the objects in to outcomes. Providing alternative codes, reaching valid decisions for the solutions of the problems are desired outcomes. During the process, learning programming concepts and solving problems were inseparable. Thus, sharing knowledge or creating ideas for enhancing basic conceptual knowledge of programming and improving structural knowledge to provide effective programming codes were the expected acquisitions. Tools are the voice, video, desktop share, survey, and the others as well as course content in web conferencing system. Most of the students experienced to use the tools in previous year. Tools were generally easy to use but when most of the
tools were used at the same time, even a little contradiction may occur on the interface. Community in this study includes students, instructors, and the staff of the organization. Norms set by the organization were assessed in Rules. The instructor also provided norms that circumscribe the activities. Division of labour was considered as the collaborative study among students for the tasks assigned by the instructor. Within each sub-component of the framework, results from observations were presented with student perceptions as follows.

**Subject-Tools-Object**

During the process, chat, desktop, and document share tools were used. Based on the interviewees’ perspectives, while the positive effects of the tools are underlined, some students also had negative views. Some of the noteworthy responses in this respect are presented.

“...Sometimes we intended to ask questions about topics that we could not understand but we avoided to interrupt the instructor” (s2).

“...The interaction between instructor and a student could occur individually that positively influenced the instructional process” (s5).

Considering the student perspectives in Subject-Tools-Object, the frequency of students’ usage of the tools during the courses pertaining to the Subject-Tools-Object component is summarized in Table 3.

<table>
<thead>
<tr>
<th>Subject-Object</th>
<th>Tools</th>
<th>Frequency of student usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student learning</td>
<td>Desktop Share</td>
<td>Frequently</td>
</tr>
<tr>
<td></td>
<td>Usage of Adobe Connect</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Video Share via Camera</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Typing in Chat Window</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Document Share</td>
<td>Rarely</td>
</tr>
</tbody>
</table>

As a requirement for the programming course’s nature, the desktop share feature of the Adobe Connect environment was used to write programming codes, and the interaction with the students was provided via related tools. In order not to disrupt flow of the course some students often asked their question at the end of the lesson. In such cases sometimes students loss of interest towards constructing knowledge. Especially, using desktop sharing following stepwise method is observed to facilitate the achieving the tasks learning process.

**Subject-Rules-Object**

Students’ with taking instructor’s permission for speaking was frequently observed. Students frequently desired to communicate either visually or orally. The themes derived from students’ perspectives are presented in Table 4.
Table 4. Results pertaining to Subject-Rules-Object component

<table>
<thead>
<tr>
<th>Subject-Object</th>
<th>Rules</th>
<th>Frequency of rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student learning</td>
<td>Unable to speak without permission</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Having the required software and hardware to use Adobe Connect</td>
<td>Frequently</td>
</tr>
<tr>
<td></td>
<td>Permission to communicate as written or oral</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Acceptance or rejection of intervention request</td>
<td>Occasionally</td>
</tr>
</tbody>
</table>

Students’ views referring to the themes emerging in the Subject-Rules-Object context are as follows:

“...Sometimes the instructor encourages us to speak, but students who do not want to speak were not involved in the environment because of the various excuses (like lack of a microphone)” (s1).
“...You can speak, share a document or present something if the instructor permits” (s3).
“...When students write irrelevant sentences on the chat window, the instructor warns...” (s5).

It is seen that students have almost necessary tools for learning programming. They can use text and oral communication in Adobe Connect with the instructor permission. It was frequently observed in nearly half of the videos. We noticed that the rules in the online courses affected pedagogical flow of the courses positively. Even it occurred a few times, the rules could restrict some students who would like to communicate with teacher and other students orally or visually. It could be difficult to explain alternative solutions or codes for tasks only using chat tools.

Subject-Division of Labour-Object

There are several views for the collaboration investigated in this study that something differs from a traditional classroom. It is remarkable that these views are frequently underlined. The frequencies are presented in Table 5.

Table 5. Results pertaining to Subject-Division of Labour-Object component

<table>
<thead>
<tr>
<th>Subject-Object</th>
<th>Division of labour</th>
<th>Frequency of rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students- Learning</td>
<td>Sharing knowledge among each other</td>
<td>Frequently</td>
</tr>
<tr>
<td></td>
<td>Following the troubles during the lessons</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Completing each other’s imperfections</td>
<td>Occasionally</td>
</tr>
<tr>
<td></td>
<td>Discussing in chat window</td>
<td>Rarely</td>
</tr>
</tbody>
</table>

We observed that collaboration among the students during programming course contributed to build new knowledge of programming and to produce effective solution for programming problems interactively. In addition, engaging in cooperative activities might enhance students’ sense of community in online courses.

Community-Tools-Object

Communication with the instructor permission, desktop share, and students’ attendance to the lectures under the supervision of the instructor were addressed within the community-centered activities. Furthermore, technical problems, document share, preparation of surveys are relatively less emphasized issues in Community-Tools-Object. Some selected expressions are presented below.

“When the instructor or staffs are absent when I need them, I feel bad about the course or the system” (s2).
“...Sometimes the system stops and do not respond (related to hardware, running programs, or high internet traffic etc.) If the instructor cannot fix it, he asks for help from staff” (s3).

Table 6. Results pertaining to Community-Object-Tool component

<table>
<thead>
<tr>
<th>Community-Object</th>
<th>Tool</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructor- Effective Instruction</td>
<td>Students’ can speak when permitted</td>
<td>Frequently</td>
</tr>
<tr>
<td></td>
<td>Supervision of students’ attendance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Desktop share</td>
<td></td>
</tr>
<tr>
<td>Organization- Effective Instruction</td>
<td>Document share</td>
<td>Rarely</td>
</tr>
<tr>
<td></td>
<td>Survey administration</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Limiting the access to some windows</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Survey administration of the staff</td>
<td>Rarely</td>
</tr>
<tr>
<td></td>
<td>Reporting technical failures to the organization</td>
<td></td>
</tr>
</tbody>
</table>
The themes pertaining to the Community-Tools-Object are presented in Table 6.

Emerged technical and connection problems during the online courses may be considered as obstacles for learning programming language. Some of the students had difficulties trying to understand the programming problem when technical problems occurred in online environment.

**Community-Rules-Object**

In this dimension, the rules that are either pre-specified or naturally created within the process for the interactions between the students and instructor were discussed. The observation data showed that, the instructor’s attendance to the system on time, providing the required equipment, giving and supervising the required permissions for the students during the classes, supervision of the messages in chat window and screen views were the prominent themes. The themes pertaining to this component were presented in Table 7.

<table>
<thead>
<tr>
<th>Community-Object</th>
<th>Rules</th>
<th>Frequency</th>
</tr>
</thead>
</table>
| Instructor- Effective Instruction | Entering the class on time  
Providing the required equipment for the lecture  
Permitting students to speak and share screen and video  
Supervision of messages on screen by the instructor  
Reporting technical failures to the organization | Frequently |

Some of the student views in the dimension of Community-Rules-Object as follows.

“...instructor continuously attempts to check whether the students were online. Students can enter whenever they want” (s1).
“...instructor enables us to attend the lecture by using participant list (s2).
“... During the lessons, the instructor instantly asks questions. He continuously engages us in the course that we are looking at the screen and paying attention to instructors’ behaviours. (s3).

We observed that the rules in the online lessons supported to improve engagement and persistence of students. Students may be focused on learning tasks in the lessons by controlling student presence by instructor and permitting sharing of students.

**Community-Division of Labour-Object**

The themes in this dimension include the interactions of the individuals in the community with each other for the purpose of obtaining information as well as providing classroom management. The observation results are presented in Table 8.

<table>
<thead>
<tr>
<th>Community</th>
<th>Division of Labour-Object</th>
<th>Frequency</th>
</tr>
</thead>
</table>
| Students-Effective Instruction | Class president acted as a moderator in the interaction between the instructor and students  
Instructor’s moderator role for the discussions of students in chat window.  
Getting technical support from the organization  
Homework share among students | Frequently | Rarely |

It was seen that instructors’ role in web conferencing system is crucial in terms of classroom management and interaction. Additionally, instructor acted as a mentor and guide for creating and managing the cooperation and interaction. The instructor tended to allow students share and communicate with other students as a facilitator to reach desired learning outcome for the tasks. Using messaging only through chat area students underlined that student-instructor and student-student interactions were somewhat restricted. On the other hand, they found instructors’ efforts for student engagement in the learning process remarkable as addressed below.
“...Instructor is giving us short-time for small examples. Then, he is allowing us share the screen and talk about the tasks. The programming course is challenging to learn and it always needs control. I need to pay more attention in the activities. The instructor can recognise me” (s1).”
“...Instructor asks whether I could do or not and adds ‘Come, let’s do it together’” (s3).

In sum, the findings of this study reflect the process in terms of organization, classroom, and participant domains. The organization domain served the affordance of tools and resources in order to achieve the tasks. The tools and activities provided students to interact each other and work collaboratively. In classroom domain, participation through the objectives was supported, and the desired objectives were positively contributed for students’ engaging in the tasks.

Discussion

The instructional process in the online programming course was discussed through Mwanza’s AT notations in this study. The results indicate that the technical features of the web conferencing system, the nature of the course and the participants’ characteristics were the key factors in the process. For example, the hardware and software resources for engaging students in the course can be evaluated in the subject-tools-object dimension. Desktop and document share in the system were discussed under the subject-tools-object component. In their study, Baran and Cagiltay (2010) revealed such situations as participants’ socialization, internalization, externalization, and their combinations by using digital video, discussion lists, and scientific publications within the framework of subject-tool-object. Similarly, various collaboration practices were provided in many of the activities during the lessons. The practices such as document sharing and the discussions in a chat window can be considered as the examples of the relationship between subject, object, and division of labour.

The interaction of the community in the web conferencing system was observed in various AT dimensions. For instance, while the tools used in the practices bearing on a large part of the community are considered in community-tool-object, a set of rules applied to engage the students in the environment were defined in the community-rules-object dimension. Instructor were played a key role for performing the rules such as speaking with permission, attending class on time, etc., under the dimension of community-rules-object. In this regard, the system provided sufficient authorization for the instructor. Similarly, Baran and Cagiltay (2010) indicated that the activities such as sharing in the community and providing technical support can be considered in community-division of labour sub-dimensions for online learning environments.

The results of this study provide detailed information about the instructional process of programming course through web conferencing system. However, it was observed that the characteristics and motivations of the subjects are sometimes required in understanding how some activities are conducted. For example, motivation and engagement of the students occur by relying on the instructor’s experience while some students need to have eye contact with the instructor for their motivation. Additionally, it was observed that some students who do not like typing never used the chat window. Similarly, there are some time restrictions about speaking in the lessons whenever the students want.

On the other hand, some researchers noticed that developing the programming knowledge is a highly complex process (Vujoševic-Janicic & Tošić, 2008). Thus, several factors might influence the quality of online programming instruction. Especially, in online programming courses, repetitive failures during lessons may lead students to lose enthusiasm and interest (Law, Lee & Yu, 2010). In this study several instructional strategies such as providing opportunities of instant feedback, collaboratively code writing, as well as discussions on codes were emerged to engage students endeavoring on the learning tasks. In parallel with this, Cakiroglu (2014) reported some evidences revealing that the quality of online programming can be supported by providing distinctive contexts. While, the nature of the programming course includes syntactic, structural, and strategic knowledge, it might be partially influential of the contexts created in the activities. The analyses within the AT sub-dimensions were somewhat relevant with the contexts.

Garrison (2007) put forward that online learning is a result of collaborative work characterized by instructional orchestration. In this study, the learning tasks were found at the center of the orchestration. Students experienced new perspectives and constructed knowledge through the tasks by interacting with others. Some direct interactions were also noticed towards various components. According to the results, the programming learning context delivered through web conferencing system is summarized in Figure 6.
According to Figure 6, since technological and pedagogical rules were discussed in the rules dimension, the technological affordances and organizational support were considered in the tools dimension. Collaboration is indirectly related to the rules. Student-student and student-instructor interactions were in the frame of Technological and Pedagogical Rules and they were assessed in division of labour. While the Learning Tasks were thought as the object, the outcomes were taken as place enhancing cognitive performance and socialization. Students, organization, and instructors are the persons were evaluated as community dimension of AT.

The findings of this study show that learning outcomes of online programming courses have affected instructional methods, web conferencing system, and role of instructor in online courses. Within the affordances of web conferencing system students’ cognitive performances were enhanced in various phases by building social connections, engaging collaboration activities and discussion, sharing their knowledge and artifacts during online courses. Even a little, some limitations were observed due to the students’ use of the tools according to their individual characteristics. Accordingly, Mwanza’s notations derived from AT provided a comprehensive framework to recognize the students and instructor in web conferencing system, cognitive processes involved in the use of tools, as well as the development of the interactions in order to reach learning outcomes.

This study has some limitations. Firstly, the video analysis was done by typical characteristics of Adobe Connect web conferencing system. Secondly, although the participants were totally 56 students, a few students were absent in some courses.

**Conclusion**

The results indicate that Mwanza’s Activity notations can be used to explore the interactions in a programming course delivered through web conferencing system. The context was interrelated as the integration of subject, tool, object, community, division of labour, and rules components. It is concluded that the instructor and students are required to have readiness at a sufficient level pertaining to either the course or the system used in the context of an online programming course. Although, some students’ individual characteristics such as interest, cooperation, responsibility, and self-efficacy need to be improved to effectively take advantage of this system in which the web conferencing system seemed to provide adequate affordance. It is an argument encountered in different sub-components of AT that the instructor’s motivation is also another factor affecting the quality of this kind of online instruction.

Students mostly have positive views about the activities to use and take advantage of the tools for the subject-tools-object component. They expected permanent solutions for problems during the desktop share, and a waste of time during the discussions. For the subject-rules-object component, many students suggested having more time for them to speak. In the subject-division of labour-object component, discussions on the chat window within learning presence were considered as positive.
In the community-tools-object component, communication with the instructor depending on permission, opportunity for desktop share and supervision of the instructor to control students’ attendance in the class is the underlined themes. Additionally, technical problems, document share, distribution of surveys by the instructor, and organization are less often emphasized. Discussions in a timely manner, provision of the required tools, instructor permissions, supervision of the messages and desktop shares in the chat environment can be considered within the community-rules-object component. The activities conducted during the lessons were generally considered as positive. Interestingly, fake students’ entrance in the activities was found somewhat different from other online learning environments. It is highlighted within community-division of labour-object component that the instructor’s role is important in classroom management and interaction. Furthermore, it is concluded that the individual differences of the students affected the outcomes of the programming course. So, while Mwanza’s sub-dimensions are used to explore the instructional processes, the nature of the course should be taken into consideration. The courses including various conceptual, operational or procedural structures may require various analysis units.

The teacher-centered activities were not so effective in depth learning; they were only effected in developing conceptual knowledge. Discussions on chat window and sharing desktop can be considered as scaffolding elements. In addition, problem-based learning activities promoted brainstorming to build agreement on the code pieces and to validate the algorithms. Thus, students developed new experiences and understandings through the learning tasks by interacting with others. Students’ responses reflect that the instructor by using the system affordances was an active member in constructing the learning community. Especially, instructors’ role was prominent, in developing strategic knowledge as developing algorithms, applying necessary programming structures for the problem. Instructor could design the online discussion tasks for fostering comprehension, exchanging knowledge (solution ways or different ideas) or enabling students for collaborative group-work. Although some technical problems occurred for short-time, it did not significantly effect to completion of learning tasks. In such cases, repetitive failures were prevented by the instructor helping students to keep them participating in productive dialogues.

**Implications for future studies**

In online courses, it is necessary to ensure that all equipment used by students function properly. Some new mechanisms need to be developed to check if the attendees are the real students. In this respect, instructors may check students’ identity by asking a set of questions. On the other hand, the inclusion of the technical team in the process is important considering the problems such as technical insufficiency in online instructional processes and the lack of timely intervention for the possible faults by the staff working in the background.

This study has considerable implications for teaching programming through web conferencing system. First, instructor should allow students connecting in live video. Second, the collaborative group activities and reflective practices should be developed by considering the individual differences, programming course content and system tools. Third, instant feedback should be provided in discussing about programming learning tasks. Overall, instructors should focus on how to design learning environments in order to enhance learner’s interaction or how to engage them in the learning tasks.

Consequently, while an instructional design is done for online courses, the designers or instructors may build an appropriate framework for people, materials, organizations, and the whole system. Personal characteristic of participants should be considered as a necessary dimension in exploring the instructional process through web conferencing systems. Thus, AT as an inclusive framework may be a comprehensive way for exploring online courses through web conferencing such as programming or other courses which have similar conceptual and operational nature.

**References**


Usefulness of Social Network Sites for Adolescents’ Development of Online Career Skills

Mariëlle Rutten1*, Anje Ros2, Marinka Kuijpers1 and Karel Kreijns1

1Welten Institute, Open University, The Netherlands // 2Fontys University of Applied Sciences, The Netherlands // m.rutten@ou.nl // a.ros@fontys.nl // mact.kuijpers@caiway.nl // k.kreijns@ou.nl

*Corresponding author

(Submitted June 25, 2015; Revised January 19, 2016; Accepted February 29, 2016)

ABSTRACT

Schools have an important role in teaching students how to use Social Network Site (SNS) for career purposes. This involves the opportunity for students to practice online career skills. Different types of digital environments are available for schools. There are SNS designed to enable users to interact and network. In addition there are digital environments – like Virtual Learning Environments (VLE’s) – designed to support learning processes but with possibilities to network and interact as well. Little is known about how characteristics of these environments may support the development of online career skills. In this study the suitability of four digital environments – two SNS and two VLE’s- to practice online career skills was investigated. The concept of affordances was used to relate the characteristics of the digital environments to how they may stimulate the practice of online career skills. Based on this concept a framework was developed to analyse the characteristics of each environment. In addition, these results were compared with the actual use of the SNS by students. The results show that, although the two SNS appear to offer a stronger environment for practicing online career skills, students do not show more online career behaviour in these environments. Implications are discussed.

Keywords
Social network sites, Online learning environments, Career learning, Online career skills, Affordances

Introduction

With the emergence of social media, the internet has transformed into an interactive environment offering multiple opportunities for communication, knowledge-sharing and development, and cooperative production. Due to this development, the internet has become an integral part of the working and learning context for all people. The internet also appears to be of growing importance for any individual who wants to direct and develop his or her career (Nikolaou, 2014; Reid, 2012). Social network sites (SNS), in particular, are used by individuals to direct their career by presenting themselves and their resumes online, by building a network, and by using it to create new opportunities for their careers (Roobol & Oonk, 2011). Since SNS are now a major context for career exploration and development (Nikolaou, 2014; Reid, 2012), individuals need to learn how to deal with this context (Gerard, 2012; Hooley, 2012).

Although young people may be more familiar with using SNS, this does not mean that they are able to use a SNS appropriately for career purposes (Benson, Morgan, & Filippaio, 2014; Gerard, 2012; Longridge & Hooley, 2012). Research has shown that students first start experimenting with how they present themselves online at age 13 (Valkenburg & Peter, 2011; Pempek, Yermolayeva, & Calvert, 2009). However, these young students are not aware of the opportunities and risks involved in using an SNS in general (Vanderhoven, Schellens, & Valcke 2013; Valkenburg & Peter, 2011), nor do they consider the possible consequences of their online behaviour for their future career (Benson et al., 2014). Therefore, different researchers suggest that schools should teach young students (age 13-16) how to use SNS in a safe, appropriate and effective way (Vanderhoven et al., 2013; Valkenburg & Peter, 2011; Subrahmanyam & Greenfield, 2008).

Students attending prevocational education often start working at age 16. Schools for prevocational education therefore increasingly recognize the importance of preparing their students to cope with changes in their professional context and ambitions. Since many of their students start using SNS at age 13, it seems logical to expect that these schools integrate the use of a SNS and the development of online career skills in their career learning programme. However, in one of the few studies on the role these schools can have in guiding young students in learning to use SNS, Vanderhoven et al. (2013) conclude that the guidance schools offer is incidental and not integrated into the curriculum. Moreover, these schools seem to be reluctant to offer students possibilities to practice their online skills by using a SNS as part of classroom practices (Vanderhoven et al., 2013; Greenhow, 2011). Schools would feel more comfortable using a Virtual Learning Environment (VLE) or a
SNS specifically designed for educational purposes (Greenhow, 2011), instead of using a SNS designed to enable its’ users to communicate with others and network. But are these VLE’s as suitable for practicing online skills as SNS? Research on how different types of digital environments may support different learning purposes is scant (Tzeng & Chen 2011; Petrovic, Jeremic, Cirovic, Radojicic, & Milenkovic, 2014). The goal of this study is to understand how the characteristics of different SNS and VLE’s -referred to as “digital environments”- may support students to practice and learn their online career skills.

**Online career skills**

Due to technological and economic developments, it is important for young people to learn to cope with the ongoing changes in professions and career perspectives. Since some students start working at age 16, many schools for (pre)vocational education no longer only provide career guidance, but have shifted their attention to career learning which focuses more on the students’ development of competencies and skills that will enable the individual to deal with changing professions and careers (Meijers, Kuijpers, & Gundy, 2013; Kuijpers, Meijers, & Gundy, 2011). Indeed, a recent study on effects of career learning shows that it is possible to teach students age 12-16 necessary career competencies (Kuijpers & Meijers, 2015).

Although different researchers define career learning in different ways, they all seem to agree that career learning is, on the one hand, a reflective process in which the individual learns about his talents, values and ambitions by giving meaning to different experiences. And on the other hand, it is an active process of formulating a possible next career step and taking steps to realise it (Kuijpers et al., 2011; Law, 2010; Goddard, 2010). Research in the past decade has shown what career competencies and skills an individual would need to develop in the process of offline career learning (Kuijpers et al., 2011; Meijers et al., 2013). Although there may be a link with offline career competencies, it seems that using a SNS for career purposes is not simply a question of translating offline skills to an online context (Nikolaou, 2014; Hooley, 2012; Gerard, 2012; Strehlke, 2010). Empirical studies show that using an SNS for career purposes seems to mainly address the active part of the process of career learning (Benson et al., 2014; Nikitkov & Sainty, 2014; Longridge & Hooley, 2012; Gerard, 2012). As for example Benson et al. (2014) show, teaching students to use a SNS for career purposes is mainly about teaching them how to present themselves online and about teaching them how to network online in a way that supports their career.

Presenting oneself can be seen as a skill related to the active process of career learning. Presenting oneself is letting others know about ones talents and ambitions. Presenting oneself online was originally done through writing a self-descriptive profile text. However, the role of the profile in SNS has changed over the years and profiles are now a more dynamic presentation of oneself in a combination of self-descriptive text and experiences (status updates) shared online (Ellison & Boyd, 2013). As the available explorative research on using a SNS for career purposes also indicates, using the possibilities to present oneself online seems to involve two different online career skills: introducing oneself by writing a self-descriptive text about one’s talents, interests and career ambitions (Benson et al., 2014; Strehlke, 2010) and enlarging ones visibility by actively sharing experiences and creating online content that represents one’s talents, interests and work experience (Hooley, 2012).

Networking can also be seen as a skill related to the active process of career learning. It is about connecting to others who may be able to help with realising a next career step. Online networking differs from offline networking. This seems to be mainly due to the fact that SNS offer the opportunity to view and traverse the connections of other members of the SNS. As a consequence, it is more transparent how ones connections may be helpful in supporting ones career or in connecting to others who may be helpful. Using these possibilities in an effective and appropriate manner is, however, not something students automatically learn (Gerard, 2012; Benson et al., 2014). Although online networking is complex (Gerard, 2012), two basic skills can be distinguished which are useful for young adolescents to start with. The first skill is connecting, which is building an online network that can support one’s career (Benson et al., 2014; Strehlke, 2010). The second is interacting, which is communicating with others to maintain visibility and strengthen one’s online relations (Nikitkov & Sainty, 2014; Gerard, 2012; Hooley, 2012).
Determining the characteristics of different digital environments

Different types of digital environments are available for schools who want to offer students the possibility to learn and practice their online skills. Different types of SNS, like Facebook or an SNS designed for educational purposes, as well as different types of VLE’s may offer the basic functionalities for students to practice the necessary online career skills. However, research on the use of digital environments in education has shown that the mere fact that digital environments like a SNS or VLE offer the functional possibilities, does not mean that these possibilities are actually used (Kreijns, 2004; Kirschner, Strijbos, Kreijns, & Beers, 2004; Kreijns, Kirschner, Jochems, & van Buuren, 2007). To examine the potential of different digital environments, a framework is needed to understand how these environments may enable students to practice their online career skills.

The concept of affordances has been used in several studies (Wopereis, Sloep, & Poortman, 2010; Tay & Allen, 2011; Wang, Woo, Quek, Yang, & Liu, 2012; Kreijns et al., 2007; Kirschner et al., 2004) to understand how the characteristics of a digital environment may stimulate learning activities. Gibson (1977) described affordances as the properties of the relationship between the physical characteristics of an object and the characteristics of an actor. This relationship enables the specific interactions between actor and object. Kirschner et al. (2004) described two principles to describe the nature of this relationship. The first principle is referred to as perception-action coupling: the moment a user becomes aware of an entity, this awareness may stimulate certain actions on the user’s part (Kirschner et al., 2004). The second principle is the principal of the reciprocal relationship between needs of the users and the possibilities offered by the digital environment to fulfil these needs (Kirschner et al., 2004).

These principles were used in this study to distinguish between two types of characteristics of the digital environments. The first category of characteristics is the trigger characteristics which are based on the principle of perception-action coupling (Kirschner et al., 2004). For example, a student notes in his digital environment that someone else shares an experience on the network. In this example the trigger characteristics are those characteristics of the environment that stimulate the student to perceive the environment as an environment in which a response to the shared experience is possible or desired. Two types of trigger characteristics can be distinguished: instructional directions given in the environment, such as for example, “about yourself,” “write your response” or “I like” and the visibility of exemplary behaviour of others, such as for example, the visibility of responses from others.

Based on the principle of the reciprocal relationship between needs of the users and the possibilities offered by the digital environment to fulfil these needs, the second category of characteristics is called the usability characteristics. In the given example, due to the trigger characteristics of the digital environment or due to other external triggers, such as for example teachers’ instructions, the student may feel the need to respond. Three types of usability characteristics can be distinguished: instructional directions given in the environment, such as for example, “about yourself,” “write your response” or “I like” and the visibility of exemplary behaviour of others, such as for example, the visibility of responses from others.

Research questions

In the previous sections we have argued the importance of guiding young students in prevocational education to use SNS for career purposes. Four online career skills that these young students would need to practice and develop were described: introducing oneself, sharing experiences, interacting and connecting. To practice these skills, different digital environments – SNS and VLE’s- are available for schools. Little is known about the suitability of these different environments for learning online career skills. Based on the concept of affordances we have argued that, although all these digital environments may have the functionalities to make practicing these skills possible, the specific characteristics of the digital environments may influence the extent to which the students may be triggered to use and will easily use these environments to practice their online career skills. The central question in this study is: What digital environment is most suitable to practice online career skills for students in prevocational education?
Method

In this study four schools, each working with a different digital environment, stimulated their students (age 12-14) in pre-vocational education to use the digital environment to practice their online career skills during a period of two months. Based on the concept of affordances an analytic framework was developed to analyse the characteristics of the digital environments used. To relate the characteristics of the four different environments to students’ online career behaviour, the profile pages of the students were analysed.

The digital environments

Four digital environments were selected for this study: two SNS and two VLE’s. The two SNS represent the digital environments that are designed to stimulate users to communicate with each other and network. The first SNS selected is Facebook, publically available and thus offering students an environment to present themselves and communicate and network with a broad range of individuals in- and outside school. The second SNS is ning.com, which is a platform that offers schools the opportunity to build their own social network environment. Ning.com is designed to stimulate interaction and networking, but offers schools the opportunity to manage the available network and thus to control the visibility of their students and the people they interact and connect with through the social network environment.

The two VLE’s represent the digital environments that are designed to support learning processes in school and offer some possibilities to see profiles of others and interact and network with others. The available network is limited since the use of the VLE is mostly limited to students and school-staff. Inviting someone outside school to become a member can only be done by the VLE-administrator. The first VLE used in this study is a Career Learning VLE and was selected because it is designed specifically to support career learning. Students have their own profile page in which they are asked to write about their talents and ambitions, the environment offers different possibilities to interact and network. The second VLE is designed to support learning processes in schools in general. This VLE was selected in this study because it is a VLE that is used by many schools in the Netherlands and therefore often is the first environment that is considered as environment to practice online career skills.

Participants

This study focused on students in prevocational education since these students often start working at a young age and the development of online career skills is important for them. Four schools for prevocational education, for whom career learning was an important aspect in their curriculum and who had the ambition to use social media as part of the career learning process, were asked to participate in this study. The schools did not want to use more than one digital environment because they expected this to lead to miscommunications in school and lack of clarity about school policy. Therefore each school represented one of the four digital environments. In Table 1 the four different schools working with these digital environments are presented. All participating students gave permission to let a researcher analyse their online behaviour in the digital environment used in their school.

<table>
<thead>
<tr>
<th>School</th>
<th>Students’ age</th>
<th>Number of students</th>
<th>% boys / girls</th>
<th>Number of teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facebook</td>
<td>A 13-14</td>
<td>11</td>
<td>36% / 64%</td>
<td>1</td>
</tr>
<tr>
<td>Ning.com</td>
<td>B 12-14</td>
<td>22</td>
<td>82% / 18%</td>
<td>1</td>
</tr>
<tr>
<td>VLE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General VLE</td>
<td>C 13-14</td>
<td>16</td>
<td>27% / 63%</td>
<td>1</td>
</tr>
<tr>
<td>Career Learning VLE</td>
<td>D 13-14</td>
<td>54</td>
<td>44% / 56%</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>103</td>
<td>50% / 50%</td>
<td>5</td>
</tr>
</tbody>
</table>

To ensure that teachers from different schools would use a similar approach in guiding their students, the teachers were instructed to use a website designed for this purpose on which instructions and examples of ways to integrate the use of the digital environment into the career learning process were given. In particular, students were stimulated to prepare a career-oriented self-descriptive text, to share experiences that related to their occupational interests and career ambitions, and to build and maintain relationships with others by connecting and interacting with others. All teachers used the digital environment over a period of two months, for one hour a week.
Instruments

Analytic framework for trigger- and usability characteristics

Based on the theoretical framework, we described trigger- and usability characteristics that can be related to each online career skill. As described in the theoretical framework we distinguished two types of trigger characteristics (instructional directions and visibility) and three types of usability characteristics (space, ease and different possibilities). For all the resulting twenty types of characteristic we described three different levels (weak, neutral and strong) in which the characteristics may be present in the digital environment. To show how different levels were distinguished and different characteristics were described, some examples are given in Table 2.

Table 2. Examples of the distinction between weak, neutral and strong scores for each type of characteristic

<table>
<thead>
<tr>
<th>Trigger characteristic</th>
<th>Online career skill</th>
<th>Strong = +</th>
<th>Neutral = ±</th>
<th>Weak = -</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructional direction</td>
<td>Introducing oneself</td>
<td>Open questions directed at information about career related talents, values and ambitions.</td>
<td>Open questions about achievements, interests or capacities that are not career related per se.</td>
<td>No directive questions asked.</td>
</tr>
<tr>
<td>Visibility</td>
<td>Sharing an experience</td>
<td>A central part of the profile page is “the wall” on which experiences shared by others are visible.</td>
<td>Experiences shared by others are only visible when visiting the relevant person’s profile page.</td>
<td>Visible to a limited number of people. The administrator determines the settings and the visibility of the experiences shared by students.</td>
</tr>
<tr>
<td>Usability characteristic</td>
<td>Online career skill</td>
<td>Strong = +</td>
<td>Neutral = ±</td>
<td>Weak = -</td>
</tr>
<tr>
<td>Space</td>
<td>Introducing oneself</td>
<td>Open space available with unlimited number of characters.</td>
<td>Open space available but limited number of characters.</td>
<td>No open space available, presenting oneself is only possible by choosing predefined descriptions.</td>
</tr>
<tr>
<td>Ease</td>
<td>Interaction</td>
<td>It is possible to respond on one’s own profile page by pushing the “write a response” button under the experience shared by someone else.</td>
<td>A response is possible by visiting the profile page of someone who shared an experience, by using the “write a response” button under the shared experience.</td>
<td>In order to respond one must go to a different part of the digital environment that offers the possibility to send a message.</td>
</tr>
<tr>
<td>Different possibilities</td>
<td>Connecting</td>
<td>A search engine is available; a search is possible on all profile information.</td>
<td>A search engine is available but a search is only possible on names.</td>
<td>There is no search engine.</td>
</tr>
</tbody>
</table>

We asked three experts to review the way in which we differentiated between the three levels and based on their comments some adjustments were made. Two researchers used the framework together to analyse the digital environments. The descriptions of characteristics, that were not clear enough to determine the presence of the characteristics, were adjusted. The researchers then used the framework to independently score the presence of each characteristic in the digital environments, in order to determine inter-rater reliability. The results showed that 95% of the characteristics were scored in the same way.
Content analysis of the online career behaviour of students

To code and analyse the content of the profile pages of the students we used a similar approach as was used in content analyses of profile pages in other studies (Greenhow & Robelia, 2009; Jones et al., 2008; Lampe, Ellison, & Steinfeld, 2007). The description of the four online career skills was used to formulate different types of online behaviour that clearly would be an expression of these skills (e.g., writing a self-descriptive text, sharing an experience, commenting on others, receiving comments from others, number of friends). We documented the frequencies of each type of online behaviour that we explicated. In addition, because many students wrote a self-descriptive text and the information given in these descriptions was diverse and sometimes extensive, we specified different types of information given (description of talents, ambitions, work- or internships experiences hobbies and interests) in these descriptions to get a more precise picture of the career relatedness of what they wrote.

Results

The characteristics of the four digital environments

Table 3 provides an overview of the results of the analysis of the trigger- and usability characteristics of the different digital environments. Assuming that, in the most optimal digital environment, all 20 characteristics would be strongly represented, the results show that none of the digital environments that were part of this study offer such an environment. Based on the overall number of “strong,” “neutral,” and “weak” scores, the digital environment of the SNS, Facebook and Ning.com offered the strongest environment to stimulate students to practice their online career skills.

<table>
<thead>
<tr>
<th></th>
<th>Social Network Sites</th>
<th>Virtual Learning Environments (VLE)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Facebook</td>
<td>Ning.com</td>
</tr>
<tr>
<td>Introducing oneself</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Trigger: Instructional directions</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Trigger: Visibility</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Usability: Space</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Usability: Ease</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Usability: Possibilities</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Sharing experiences</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Trigger: Instructional directions</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Trigger: Visibility</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Usability: Space</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Usability: Ease</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Usability: Possibilities</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Interacting</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Trigger: Instructional directions</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Trigger: Visibility</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Usability: Space</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Usability: Ease</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Usability: Possibilities</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Connecting</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Trigger: Instructional directions</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Trigger: Visibility</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Usability: Space</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Usability: Ease</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Usability: Possibilities</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Total number</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>5</td>
</tr>
</tbody>
</table>

The characteristics of Facebook and Ning.com differed on several aspects. The first difference relates to the trigger characteristics of introducing oneself. In the environment created on Ning.com, career relevant questions (e.g., what internship- or work experiences do you have?) were added to the questions asked when creating a profile, whereas on Facebook no career-specific questions were asked when creating a profile.
Secondly, interaction between members of the digital environment had a less central position in Ning.com, in the sense that interaction was only visible when one was addressed in the information shared or when visiting the profile page of the person who shared the information. Thirdly, Facebook and Ning.com differ with respect to the possibilities (usability characteristics) offered to build a network. Ning.com offered a search engine to find people to add to your network, but a search was only possible on a name, whereas in the search engine offered by Facebook, a search on different terms is possible.

Furthermore, the results showed that the trigger- and usability characteristics of the two VLEs were weak in comparison to those of Facebook and Ning.com. The career learning VLE distinguished itself from the other VLE in a positive sense by offering a relatively strong environment for introducing oneself. This was mainly due to the directive career related questions (about future ambitions and work- and internship experiences) that were asked when a profile is created. The career learning VLE also offered a relatively strong environment for connecting in giving some instructional directions that pointed out that it was possible to connect and it offered some easy possibilities to do so. Students working with the VLE only had the possibility to connect with others by asking the administrator to give someone else access to their profile.

The online career behavior of the students

Table 4 provides an overview of the quantitative measures of the online career behavior of students in the different digital environments.

<table>
<thead>
<tr>
<th></th>
<th>Facebook</th>
<th>Ning</th>
<th>Career Learning VLE</th>
<th>VLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introducing oneself</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writing a self-descriptive text</td>
<td>0%</td>
<td>45%</td>
<td>46%</td>
<td>94%</td>
</tr>
<tr>
<td>• about talents</td>
<td>-</td>
<td>0%</td>
<td>36%</td>
<td>0%</td>
</tr>
<tr>
<td>• about preferred future profession or ambition</td>
<td>-</td>
<td>30%</td>
<td>84%</td>
<td>73%</td>
</tr>
<tr>
<td>• naming past work- or internship experiences</td>
<td>-</td>
<td>30%</td>
<td>20%</td>
<td>40%</td>
</tr>
<tr>
<td>• about hobbies and interests</td>
<td>-</td>
<td>90%</td>
<td>88%</td>
<td>53%</td>
</tr>
<tr>
<td>Sharing an experience</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of students sharing current activities and experiences</td>
<td>27%</td>
<td>0%</td>
<td>100%</td>
<td>25%</td>
</tr>
<tr>
<td>Average number of experiences shared</td>
<td>1.3</td>
<td>-</td>
<td>9.5</td>
<td>1</td>
</tr>
<tr>
<td>Range</td>
<td>0-8</td>
<td>-</td>
<td>1-10</td>
<td>0-1</td>
</tr>
<tr>
<td>Interacting</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of students that contributed to interaction</td>
<td>55%</td>
<td>32%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Average number of contributions of the students that interacted</td>
<td>6.33</td>
<td>1.86</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Range</td>
<td>1-12</td>
<td>1-5</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Connecting</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of students that connected to others</td>
<td>100%</td>
<td>50%</td>
<td>24%</td>
<td>0%</td>
</tr>
<tr>
<td>Average number of connections</td>
<td>22</td>
<td>2</td>
<td>1.47</td>
<td>-</td>
</tr>
<tr>
<td>Range</td>
<td>7-57</td>
<td>0-5</td>
<td>1-17</td>
<td>-</td>
</tr>
</tbody>
</table>

**Introducing oneself**

Students using Ning.com, the VLE or the career learning VLE wrote a self-descriptive text in their profile. Only students working with the career learning VLE wrote about their talents. These students simply named their talents: "I am creative and open to other people." The text also indicated that three students asked someone else to write a text for them, e.g., "a sweet and nice girl, also very helpful. XX B."

The majority of students working with the VLE or the career learning VLE wrote about their future ambitions and about their interests and hobbies. Only 30% of the students working with Ning.com wrote about future ambitions. Most students working with Ning.com wrote about hobbies and interests, however, from a career perspective, this information is less relevant than information about talents, future ambitions or work experience.
Connecting the findings in Table 4 to our analysis of the characteristics of the digital environments (Table 3) does not show a clear relationship. As Table 4 shows, none of the 11 students working with Facebook used the available space and possibilities to present themselves. In comparison, 15 of the 16 students working with the VLE wrote a self-descriptive text. Moreover, students using the VLE or the career learning VLE seemed to be more likely to give career relevant information in their self-descriptive text than students working with Ning.com.

Sharing experiences

Sharing experiences is about sharing experiences and creating online content that represents one’s talents, interests and work experience. Students sharing an experience mostly gave limited information and did not specify how these experiences related to their talents and ambitions. For example, a student working with Facebook shared a picture of something she had made with an accompanying note “Made this during art-class.” Other students gave more information, but a connection with how this represented their talents or ambitions was still difficult to make, e.g., a student using the VLE: “I did a one-day internship with a hairdresser. They said I could come back for a longer internship. I behaved well, worked independently and did a lot of different things even though they were not very busy that day.”

When relating the characteristics of the digital environments to the online career behavior of the students, the high amount of experiences shared by students using the career learning VLE is remarkable, especially in comparison to the students using Ning.com, who did not share any experiences. The fact that students working with the career learning VLE got several mandatory assignments to share an experience, may explain this. However, this does not explain why none of the students working with Ning.com shared an experience, nor does it explain why a relatively high amount of students (25%) working with the VLE shared an experience. Again, the relationship between the online career behavior of students and the characteristics of the digital environment they used is not clear.

Interacting

The results of the analysis of digital environments presented in Table 3 showed that Facebook and Ning.com offered a strong environment for interaction in comparison to the environment offered by the two VLEs. Table 4 shows that only students who used Facebook and Ning.com interacted online. It appears that students using a digital environment designed to interact, actually show more interaction activities online, although the career relevance of the student interactions in these environments is questionable.

The online interaction on both sites was informal and mostly a combination of one or more “likes” and comments like “oh, that looks really funny” or “tell me more!” The average number of interactions of each student shows that there was some interaction but there were no extensive conversations or discussions about a topic.

Connecting

Table 4 shows that students working with Facebook and Ning.com showed more connecting behavior than those working with the VLE or career learning VLE. The characteristics of the digital environment may have played a role in this. As the results presented in Table 3 show, both SNS used in this study offer a stronger environment for connecting than the two VLEs used by students in this study. Students working with the VLE were able to involve others in what they were doing, by asking the administrator to make them a member and give them rights to view their profile. This is a rather complicated way to connect, which might explain why none of the students used it. Students working with the career learning VLE could connect with others by making them their “favourites,” but users of the career learning VLE were not explicitly invited to add others to their favourite-list, which may explain why these students made fewer connections.

Conclusion

The increased use of SNS has created a new environment for career exploration and development (Reid, 2012; Nikolaou, 2014). Adolescents start using SNS at age 13, but are not aware of the opportunities and risks involved
in using a SNS as for example possible consequences of their online behavior for later career possibilities (Valkenburg & Peter, 2011; Pempek et al., 2009). Recent studies also showed that, even though adolescents may be active users of SNS, they do not automatically learn how to use a SNS for career purposes (Longridge & Hooijen, 2012; Gerard, 2012; Benson et al., 2014). Based on the results of these studies four types of online career skills were ascertained: introducing oneself, sharing experiences, interacting and connecting.

To develop online career skills schools should offer students an appropriate digital environment to practice their skills. This is important not only because students are active users of SNS from age 13 on, but also because students in vocational education often start working before their twenties. Developing the necessary online career competencies and skills at a young age may help them to deal with ongoing changes in professions and career perspectives when they start working. Different types of digital environments –SNS and VLE’s- are available to use in school. However, little is still known about how the different characteristics of these environments may support learning (Tzeng & Chen, 2011; Petrovic et al., 2014).

This explorative study was the first study to investigate the suitability of different digital environments to practice online career skills for students in prevocational education. Based on two principles derived from the concept of affordances (Gibson, 1977; Kirschner et al., 2004) a new analytic framework was developed for the analysis of digital environments. This framework allows us to compare the characteristics of the digital environments that may enhance the students’ practice of online career skills. Although further research is needed to validate the framework, the developed analytic framework seems to be appropriate to determine the extent to which a digital environment may enable students to practice their online career skills. Although further research is needed to validate the framework, the developed analytic framework seems to be appropriate to determine the extent to which a digital environment may enable students to practice their online career skills.

The results of the analysis of four digital environments –two SNS and two VLE’s- show that there are major differences in the extent to which the characteristics of these environments may stimulate students to practice their online skills. The characteristics of the two SNS used in this study seem to offer a stronger environment then the two VLEs for practicing online career skills. However, the analysis of student behavior in the different environments showed that students do not show more online career behavior in these two SNS. Although, students interacted with others and were connected to others to a greater extent when working with one of the SNS, the interaction they showed was mainly informal and not career related. Moreover, in comparison to the students working with one of the VLE’s, students working with the SNS did not introduce themselves more or share more experiences.

These results can be seen as a confirmation of the findings of Petrovic et al. (2014) study in which Moodle (a VLE) and Facebook were compared as learning environments for university students. Their results also indicated that students would prefer to use a VLE when giving more personal information, which they are asked to do when practicing the online career skills “introducing oneself” and “sharing experiences.” Publicly available SNS, like Facebook, seem to be more suitable for connecting and informal social interaction with others (Petrovic et al., 2014). Further research is needed to see if these results are also confirmed when using more different digital environments. The results of our study may also indicate that other factors, besides the characteristics of the digital environment, influence students’ online behavior. One possible factor is the instruction of the teacher and the assignments students were given. The teachers were asked to encourage students to act out their online career skills and to structure their lessons in a similar way. However, as the results showed, the mandatory character of the students’ assignments, when using the career learning VLE, clearly stimulated them to exhibit more online career behavior. Similarly, the way in which the teachers may have stimulated their students to practice their online career skills, by giving mandatory assignments or by rewarding them in some way, may have influenced students’ behavior as well. Moreover, research has already shown that the quality of the guidance of the teacher is crucial for the development of career competencies (Kuijpers et al., 2011; Meijers et al., 2013). Further research is needed to gain a better understanding of the role of the teacher when guiding students in the use of a SNS for career purposes, the role of more binding assignments, and the way in which these factors intervene with the relationship between the characteristics of the digital environment and students’ online career actions.

With this explorative study we made an initial attempt to gain a better understanding of which digital environments are most appropriate to stimulate students to practice their online career skills. This study has its limitations. Using a social network site for career purposes and guiding students in how to do this is a relatively new focus area for schools. Due to the limited number of schools actually using a digital environment for this purpose and due to the fact that we wanted to compare different digital environments, we were not able to randomly select schools. Moreover, the number of students working with each environment was limited as
indicated in Table 1. Although this explorative study has its limitations, it demonstrates the fact that the characteristics of digital environments may influence students’ online behavior, but that the relation is complex. Further research is needed to gain a better understanding of how different digital environments can be used effectively in schools to teach students how to use a SNS for career purposes.

References


A Mobile Application to Improve Learning Performance of Dyslexic Children with Writing Difficulties

Rabbia Tariq* and Seemab Latif

Department of Computer Software Engineering, National University of Sciences and Technology (NUST), Islamabad, Pakistan // rabiatariq.mscs19@students.mcs.edu.pk // seemab@mcs.edu.pk

*Corresponding author

(Submitted June 30, 2015; Revised February 23, 2016; Accepted March 25, 2016)

ABSTRACT

A neurological learning disability, termed as Dyslexia, is characterized by difficulties in various aspects of writing skills making the individuals unable to develop age-appropriate and ability-appropriate functional skills. In Pakistan, lack of dyslexia awareness and remedial education training restrains the remediation of dyslexic children at early stages. There also exist noticeable affordance and accessibility issues concerning the remedial help and assistive technology adoption. In this research, we have developed a mobile learning application for android-powered devices that targets developmental progression and usability requirements of dyslexic children with writing difficulties. Our center of interest was to improve introductory writing skills of dyslexic children along with consistent evaluation of their learning performance to highlight the weak areas of learning process. To achieve this objective, we have designed a Writers Learning Algorithm (WLA) based on computational model of learning. We assessed the effectiveness and usability of the developed application in collaboration with dyslexic students as well as remedial teachers at selected dyslexia institutions in Pakistan. The preliminary assessment results concluded that application has justified its role in terms of representation, evaluation and optimization of writing proficiency of young dyslexic writers.

Keywords

Dyslexia, Writing difficulties, Assistive learning application, Technological usability concerns, Learning performance measurement

Introduction

According to Pirani and Sasikumar (2012), learning disabilities are normally hidden disabilities that affect many individuals with average or above average intelligence throughout their life. Edyburn (2006) stated that the cognitive characteristics (intellectual ability, attention deficits, memory, and thinking skills) and academic characteristics (reading, writing, language arts, and mathematics) of a person are associated with various types of learning disabilities.

Importance of handwriting skill for Dyslexics

Lam et al. (2011) described that insufficient handwriting skill is considered as one of the functional deficits in dyslexia. Martínez-Marrero and Estrada-Hernández (2008) indicated that dyslexic children often experience difficulties in mechanical aspects (letter formation, capitalization, spelling and punctuation) as well as contextual aspects (organization and consistency) of writing. Edyburn (2006) remarked that it is more fruitful to overcome the writing difficulties of dyslexic individuals by adopting assistive technology at primary level as it will make the learning tasks easier throughout their academic life. Evmenova and King-Sears (2013) also emphasized that monitoring the impact of technology on individual’s writing progress is very much essential.

Dyslexia research in Pakistan

As a developing country, Pakistan faces several challenges in dealing with dyslexia as government help is not enough to combat this on a country wide scale. Besides that, the adoption of specialized learning aids is still in its infancy. Until now, no significant contribution has been made in the research area of the assistive technology for dyslexia. In order to improve the current situation of dyslexia in Pakistan, this research endeavors to design and implement an interactive technological learning aid specifically for dyslexic children with handwriting difficulties. It also addresses the prominent awareness, affordance and accessibility issues regarding the remedial educational help and assistive technology adoption.
Literature study and discussion

A detailed background study about issues faced by dyslexic individuals has been conducted to find out the usage of technological assistance in learning and improving writing skills in general, and associated usability issues in particular.

Major barriers faced by Dyslexic children

According to Pirani and Sasikumar (2012), dyslexic individuals are characterized by slow conceptual development, poor retention span and/or weak sense of spatial awareness. Ndombo et al. (2013) emphasized reading and writing as the main barriers for dyslexics. Pareto (2012) added that the difficulties in mathematical skills are even a bigger challenge than the influence of poor reading or writing skills. Flora (2009) remarked that dyslexia is commonly accompanied by Dysgraphia, a specific learning disability which causes the individuals to write words backwards, inconsistent letter formation and mix uppercase and lowercase letters. Pirani and Sasikumar (2012) indicated that dyslexic children often feel frustrated or depressed having a poor self-image which leads to failure in their academic life.

Multimedia approaches for Dyslexics

According to Ndombo et al. (2013), education of individuals with learning disabilities (dyslexia in particular) is a time consuming job that demands proper attention and extra efforts from instructors. Nicolson and Fawcett (2008) mentioned that good teaching practices for dyslexic individuals involve dividing the complex skills into sub-skills. So that individuals master the sub-skills with multiple trails on daily basis before using them as building blocks for the next skill. Mahidin et al. (2011) claimed that multimedia applications have a potential to effectively teach dyslexics children allowing more than one sensory channel to acquire and process the instructions. Wah (2007) noted that dyslexic children require repetitive instructions, thus, multimedia resources can be used to gain mastery in various learning skills as it reinforces and motivates them without losing patience. Ahmad, Jinon, and Rosmani (2013) concluded that use of multimedia applications can get individuals’ attention, improve their understanding and boost their confidence.

Existing technological learning solutions

According to Feder and Majnemer (2007), various studies of handwriting intervention predicted that most assistive tools improve the legibility of handwriting and writing problems can disappear with appropriate intervention. On the other hand, Pirani and Sasikumar (2012) argued that assistive technology cannot be generalized for individuals with learning disabilities as every learning disability has its own unique requirements. Martinez-Marrero and Estrada-Hernández (2008) also mentioned that the instructional design of the technology should be learner-centered with specific learning disability. In past, researches have been conducted to develop assistive learning solutions specifically for individuals with dyslexia. These include MyLexic (Devaraju et al., 2007), JollyMate (Khakhar&Madhvanath, 2010), E-Z Dislexia (Mahidin et al., 2011) and MathLexic (Ahmad et al., 2013).

All these researches have issues/limitations of individual performance measurement, courseware structure, technical implementation, usability and user acceptance.

Weaknesses in literature

Reviewing the relevant literature clarified that technological interventions can facilitate dyslexic children with writing difficulties as there is no cure for dyslexia until now. It also highlighted some gaps in the area of research related to dyslexia. The major concern is that the majority of existing assistive learning aids are designed to address either the reading difficulties or the writing difficulties but not both in one system. In fact, little attention has been given to solve the writing problems of dyslexic children though it is a skill required for survival of primary school children. Many of them have completely neglected the importance of handwriting technique and have provided alternative tools. In addition, the research on mathematical disabilities is very limited.
Based on the literature surveys, one of the reasons behind the inadequate adoption of assistive technology is the lack of teacher/parents understandings regarding the technology use as many of them are too difficult to operate. A large range of available learning systems for dyslexics are not designed keeping in mind both the educational goals and diverse individual’s needs making it difficult to choose the appropriate learning system for specific age group, grade level or ability level. Also the available technological solutions do not offer any performance measurement feature. In addition, there is a lack of explicit instructional design in the existing technological learning aids as dyslexic children somehow require writing assistance while interacting with the system.

**Research questions and scope**

The literature survey we conducted highlighted some research gaps in the area of assistive technology for dyslexia that motivated us to conduct this study. Keeping in mind the potential cost-effectiveness feature of mobile applications, the primary goal of the research constitutes of two research questions: Whether an application, designed specifically for dyslexic children, can foster learning and help to improve some of their fundamental handwriting skills in English language and mathematics by performing the assessment of their writing artifacts (Q1) and which design elements and courseware features of an application are important for appropriateness and appeal to dyslexic children in order to potentially improve their learning experience and decrease their learning differences (Q2). These factors are primarily measured as the “usability of the application.”

Analysis of available literature highlighted that the application framework should be designed in accordance with the age, ability and usability requirements of targeted dyslexic population. Therefore the users of proposed system are preschool dyslexic children aged 5 and under with writing difficulties. The writing activities incorporated are based on their educational goals. The environment can be both home and remedial educational institutes. The technology used to implement the system is Android software platform, compatible with both smartphones and tablets. The graphical user interface of application is entirely based on dyslexic users’ usability requirements.

**Dyslexics usability requirements for application**

**Research methodology**

As the application is tailored to the needs of the preschool dyslexic children, the software engineering methodology employed for the research is Usability Engineering Process Model (Perdrix et al., 2003). It supports the usability considerations of the targeted population throughout the development to craft the interaction designs through iterative cycles.

![Usability engineering process model](image)

*Figure 1. Usability engineering process model (Perdrix et al., 2003)*
As shown in Figure 1, each cycle comprises of four main processes:

- **Analysis:** It gathers as well as structures the user requirements with the help of literature survey and interviews.
- **Design:** It devises the application architectural model, technological framework and proposed learning algorithm concepts.
- **Implementation:** It transforms specifications into executable application using software.
- **Launch & Evaluation:** It evaluates the developed application within the learning environment to determine its effectiveness in learning the handwriting skill.

The recurring analysis, design and evaluation cycles helped greatly to progress various elements of the application during the implementation process.

**Requirements elicitation**

The design and development of the application occurred in collaboration with dyslexic children, their parents and remedial education teachers specialized in the field of dyslexia from IDEAS (Institute of Dyslexia Education & Attitudinal Studies), Islamabad, Pakistan and CARE (Center for Assessment and Remedial Education), Karachi, Pakistan. Both face-to-face and phone interviews were conducted with ten remedial teachers who shared a great deal of usability design elements appropriate for designing a handwriting learning application for preschool dyslexics. We also surveyed five parents with dyslexic children to document their opinions and supportive information for a better understanding of the everyday environment in which dyslexics perform and the difficulties they encounter.

Interviewing the remedial education teachers revealed that majority of parents refer to specialized remedial institutes in order to diagnose their child for any learning disability after they observe that their child’s struggle continues regardless of changing the teaching strategy or teacher. It is difficult to tell effectively whether a kindergarten or first grade child is experiencing a developmental delay or is dyslexic as every child start learning at slightly different age. But when a child starts to learn, he/she accomplishes the developmental milestones in a reasonable time. To establish a learning profile of child, remedial teachers evaluate age-appropriate subjective (motivation, listening, comprehension, attention and oral communication) and objective (working memory, reading fluency, decoding and word retrieval) observations using research-backed dyslexia screening tools in order to determine a good starting place for a child that needs an intervention.

**Requirements specification and analysis**

One of the aims of the research was to individualize the application for each student’s learning level so that the diverse needs of the dyslexic children are satisfied. A Usability Requirement Questionnaire was developed to conduct a survey. It was targeted to specialized teachers of dyslexia and parents having dyslexic children aged 5 and under. The questionnaire was divided into three categories namely Interface, Content and Evaluation. It consisted of carefully chosen questions; based on literature study, to gather the response data related to the following areas: level of dyslexia severity, aspects of dyslexia the child struggled with, which tasks children have difficulties with, how their confidence is stimulated, which factors engage the children’s attention, what are their treatment methods, how their learning performance is measured and their preferences regarding mobile technology features.

The analysis of survey results affirmed that the most significant part of any application is the user interface. It is supposed to deliver the feeling of comfort and satisfaction to the users while using the application. Table 1 shows the preferred user interface design guidelines to develop an application for dyslexic’s learning.

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Questions</th>
<th>User preference</th>
<th>Application specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Do you think there is any alternative that can help the children with dyslexia learn to write instead of pencil/pen?</td>
<td>Fingers as writing tool</td>
<td>Use of fingers to trace and free-hand writing the characters</td>
</tr>
<tr>
<td>2</td>
<td>Which font color do the children with dyslexia prefer?</td>
<td>Darker Shades</td>
<td>Black</td>
</tr>
</tbody>
</table>
3. Which font style do the children with dyslexia prefer? Sans serif font group Arial
4. Which font size do the children with dyslexia prefer? Neither too large nor too small 14/16
5. What background colors do the children with dyslexia prefer? Lighter Shades White, lighter shades of pink and green.
7. Which factors distract the attention of the children with dyslexia? Use of too much objects/text No unnecessary images, animation and text, Fewer options to choose from.
8. Which age group has a better learning factor? The younger, The better Made for ages 5 and under.
9. What kind of technological device dyslexic children prefer most? Tablets, Smartphones Based on android platform that can be visualize on both tablet and smartphone devices.

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Questions</th>
<th>User preference</th>
<th>Application specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>What are the prominent weakness areas of the children with dyslexia?</td>
<td>Reading, Writing, Phonemic Awareness, Penmanship</td>
<td>Phonemic awareness→character-sound knowledge. Reading→character stroke formation order with associated character sound. Writing→character Tracing and Copying worksheets. Penmanship→stylus pen.</td>
</tr>
<tr>
<td>2</td>
<td>Which handwriting style do you prefer that improves their memory and speed?</td>
<td>Print writing</td>
<td>Print writing.</td>
</tr>
<tr>
<td>3</td>
<td>What are contents of syllabus involved for the children with dyslexia?</td>
<td>Depends on age group &amp; weaknesses</td>
<td>Activities for learning basic building blocks of writing.</td>
</tr>
<tr>
<td>4</td>
<td>What kind of help/guidance is required by children with dyslexia?</td>
<td>Depends on type of learner</td>
<td>Multisensory writing instructions to accommodate visual, auditory and tactile learners.</td>
</tr>
<tr>
<td>5</td>
<td>What problems are faced by dyslexic children while writing?</td>
<td>Time management, Memory pressure, Staying focus, Illegible writing</td>
<td>Time management→Total time taken for each activity. Memory pressure→Fewer options available to choose from. Staying focus→Use of attractive graphics. Illegible writing→Rigorous revision. Learning method based on useful aspects of Scaffolding, Orton-Gillingham and Davis method.</td>
</tr>
<tr>
<td>6</td>
<td>Do you follow any particular treatment program for the children with dyslexia?</td>
<td>Scaffolding, Orton-Gillingham, Hickey, Fernald, Davis</td>
<td></td>
</tr>
</tbody>
</table>

Investigating the gathered requirements revealed that the contents of the application must be in accordance with the age and ability level of dyslexic children to prove to be beneficial. The suitable learning material design guidelines and their availability in the developed application are given in Table 2.

Analysis of the surveyed requirements concluded that the ultimate goal of evaluation is to better understand the strengths and weaknesses of dyslexic child’s performance during the learning process while acknowledging the learning efforts to improve learning. Table 3 represents the desired learning evaluation design guidelines and their availability in the developed application.
Table 3. Learning evaluation requirements

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Questions</th>
<th>User preference</th>
<th>Application specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Do the evaluation tasks given to the children with dyslexia are predefined?</td>
<td>Yes</td>
<td>Upper-Lower case letters, Mathematical numbers &amp; symbols test activities for revision.</td>
</tr>
<tr>
<td>2</td>
<td>What are the types of evaluation tasks?</td>
<td>Based on individual’s weak areas</td>
<td>Frequent assessments</td>
</tr>
<tr>
<td>3</td>
<td>What is the frequency of evaluation?</td>
<td>After 6 months</td>
<td>Frequent assessments</td>
</tr>
<tr>
<td>4</td>
<td>In what way the results of evaluation are computed?</td>
<td>Calculation of %score based on the answers</td>
<td>Display learning graph, success stars, % progress &amp; time tracker.</td>
</tr>
<tr>
<td>5</td>
<td>To whom the results of evaluation are notified?</td>
<td>To parents, To teachers</td>
<td>History section to record evaluation results of activities.</td>
</tr>
<tr>
<td>6</td>
<td>How do you represent the results of evaluation?</td>
<td>Score sheet</td>
<td>Statistics tables and Observational notes.</td>
</tr>
<tr>
<td>7</td>
<td>What is the way of appreciation/reward system followed?</td>
<td>Any type of token of appreciation and motivation</td>
<td>Token system→Display Success Stars. Audio feedback in the form of applause sound.</td>
</tr>
</tbody>
</table>

Application description

Both the literature survey as well as interviews with remedial education teachers and parents of dyslexic children highlighted some learning barriers that dyslexic children generally struggle with. Hence such individuals require a different approach in learning. They also mentioned that applications designed for mobile devices will have a significant impact on learning process of children with writing differences. Thus this mobile application is specifically developed based on the understanding of characteristics and preferences of dyslexic children aged 5 and under and its effects on learning.

Architectural model

The architectural model of the application is comprised of various lesson plans and tests activities as depicted in Figure 2. The particular content structure is selected based on the existing remedial education methods tied with the diverse needs of dyslexic children to remediate their learning barriers. Also, specific content representation techniques based on literature survey and input from the instructors/parents of dyslexic children are implemented to provide assistance in their learning process. The application is structured around three basic modules where each module is designed with specific user interface to achieve a specific learning goal:

- **Learn Module**: This module is designed to strengthen dyslexic children’s reading and writing skill while fortifying their time management skill. Also, this module is believed to improve their phonological processing through character-sound association. In addition, the representation of character stroke formation order eases dyslexic child’s difficulty in following the writing instructions. Each category (upper case letters, lower case letters, mathematical numbers and symbols) within the module comprised of different lesson plans, based on multisensory approach, offers each child the facility to attempt and proceed to the next lesson plan based on the needs and learning performance.

- **Test Module**: Aimed at improving dyslexic child’s visual memory. Test activities require dyslexic child to recognize, memorize and identify visual character shapes in order to rewrite them. This helps the child to remember and recall the gained knowledge in an easier and quicker way.

- **Evaluation Module**: A unique element designed to perform the assessment of the writing worksheets. The scoring system implemented is very moderate, in order to boost dyslexic children’s confidence to enhance their learning performance. Also, it motivates them to increase their involvement with the application by appreciating their learning efforts.
Two styles of writing activities i.e., trace mode and copy mode are incorporated in each module. Within Learn module, the child is provided with writing assistance in form of simple instructions. In order to toughen the writing activities, this feature is not available in Test module. Besides this application comprises of additional modules for example documenting the data related to improvements in child’s learning outcomes:

- History module: Intended to function in collaboration with evaluation module demonstrating the structured records of child’s learning growth in the form of progress tables.
- Workbook module: Designed to record the observation notes related to child’s learning behavior providing a way for teachers/parents to view the improvements in child’s learning attitude.

![Handwriting Application](image)

Figure 2. Architectural model of application

**Technological framework**

The technological framework of the application as shown in Figure 3 is designed to eliminate the following learning barriers in one system simultaneously.

**Phonological barriers:** The application used the audiovisual elements to make a child familiar with units of character sound called phonemes. Once the learner gained the skill of phonological awareness, he/she will naturally understand how to read.

**Reading barriers:** The application used the visualization concepts to help a child memorize the character effortlessly. Therefore, after acquiring the skills of phonological awareness and reading, the learner will be able to write.

**Writing barriers:** The application implemented the custom designed learning algorithm to evaluate the child’s writing performance and to notify the learning memory aptitude.

The activities incorporated in application are designed to impart the knowledge of phonology, reading and writing. The application is expected to progress the writing skills of dyslexic children by implementing proposed
learning algorithm to make dyslexic child aware of his learning growth by identifying the weak writing areas where further improvements are required.

Prototype demonstration

As the application manages multi user profiles, therefore each user account possess a unique login ID and password. This allows secure sharing of one device among many users while keeping the separate records of individual’s learning performance data. Figure 4 demonstrates the screenshots of the representations related to various modules of developed application.

Proposed learning algorithm concepts

Research study and interview sessions highlighted some learning concerns relevant to dyslexic students learning process that are mainly associated with illustration of courseware contents, determining the learning progress and testing knowledge set hence requiring a custom designed algorithm, based on computational model of learning. The concepts incorporated in the Writers Learning Algorithm (WLA) are believed to solve all these learning differences by providing dyslexia friendly learning environment. Figure 5 shows the devised WLA which
consists of following three components to remediate the prominent learning barriers. More specifically, the detail of each component is presented below:

**Representation**

This component is aimed at improving dyslexic child’s comprehension and recognition skills by implementing the re-learning mechanism that intends to repeat the same character set until it is learnt properly. The reason behind re-learning is that these individuals are regarded as slow learners and they can take a week or even months to learn single character. The WLA presents dyslexic child with single character set at a time that requires him/her to recognize, memorize, recall and write the character without any confusion. The alphabets, numbers and symbols are divided into enumerated character sets \((n = 1, 2, 3...)\) where number of elements within each character set varies from section to section. The WLA selects a character set from specific section for purpose of representing it to learner based on individual’s needs and learning capacity. It then evaluates each element of selected character set \((n)\) within both writing modes to decide whether to transit to next mode or to repeat the current mode of selected character set \((n)\). Also it decides whether to move to the next character set or not depending on the evaluation result. As there is finite number of pre-defined character sets, the WLA ensures that activities keep flowing in a recurring manner without getting stuck.

*Figure 4. Prototype demonstration*
Evaluation

Although it is quite cumbersome to observe day to day learning progress of dyslexic child, the goal of this component is to accomplish the frequent assessments on the writing worksheets in order to judge the learning memory. Also it is designed to motivate the child to increase the learning efforts by displaying the learning growth while simultaneously increase the possibilities of achieving the optimal learning performance result. The WLA measures and records the child’s writing performance for each module to determine how well he has performed. An Evaluation function is used by the WLA to evaluate each element of a character set \((n)\) within both modes based on the performance criteria:

\[
\text{Character \%score} = \left( \frac{\text{Obtained points}}{\text{Total points}} \right) \times 100
\]

The WLA uses a Scoring function to compute the learning performance against a character set \((n)\) within both modes:

\[
\text{Average \% Score} = \left( \frac{\text{Sum of character \%score}}{\text{Total no. of Set Elements}} \right) \times 100
\]

The WLA also exhibit the learning rate of dyslexic child by tracking the time taken to accomplish learning a character set \((n)\). It determines the learning growth of a child with the help of graphical representation of evaluation results showing a character as a function of time (no. of attempts) and learning proficiency in each mode (average \%score).

Optimization

This component aims at providing the rigorous practice experience to test dyslexic child’s knowledge set because they have significant difficulty in retaining information in memory. It is designed to revise what has been learnt.
yesterday or before in a suitable way to over learn the skills already mastered. Thus it is expected to greatly help dyslexic child with remembering and recalling the gained knowledge without any memory pressure. The WLA is intended to optimize the knowledge set of dyslexic children by implementing reinforcement learning mechanism. Each writing activity is based on the principle that the more one does something, the better one gets at it. This develops self-confidence in dyslexic child to accomplish writing tasks which he has learnt before. In addition, it allows the self-motivated learning by frequently appreciating his learning efforts.

Evaluation and result discussion

After the design and implementation process, we collaborated with the IDEAS (Institute of Dyslexia Education & Attitudinal Studies), Islamabad, Pakistan and CARE (Center for Assessment and Remedial Education), Karachi, Pakistan to evaluate the developed application. The research used the quantitative evaluation approach by conducting the usability and user acceptance testing in order to appraise the effectiveness and usage of the application. The main purpose for evaluating the application was to identify the important learning, design, conceptual and technological issues during the supervised interaction of dyslexic children with the application.

Participants

Twenty dyslexic students aged 5 and under were recruited from selected remedial education institutions using convenient sampling. They used English language as their secondary means of verbal and written communication and were right hand dominant. The following information was obtained from the remedial education teachers regarding the participants: 30% participants (age 3-4 years ± 06 months) were having problems in reading where more than half were diagnosed with phonemic difficulties too. 70% participants (age 4-5 years ± 06 months) were receiving assistance in handwriting with larger part registered as having significant problems in penmanship. Children with cognitive deficits or developmental delay were excluded from this study as such conditions might interfere with other abilities. The parents of all participants gave their informed consent.

Evaluation criteria

During the evaluation process, a User Acceptance Test Questionnaire was developed which comprises of structured sets of questions, carefully chosen from literature survey, to reveal the improvements in the writing skills of dyslexic children. Questionnaire was divided into four categories i.e., Interface, Content, Evaluation and Appreciation. The questionnaire was based on Likert Scale with numbered responses marked from 1 (strongly disagree) to 5 (strongly agree) measuring the compliance of the various key factors of usability and user acceptance in the application.

Measuring system/user performance experience

As the application design was believed to be learner-centered, so it was essential to find out whether the features integrated in application do actually support dyslexic child’s learning process and assist in enhancing the handwriting skill. To do so, we evaluated the usability of the application within the learning environment by employing the well-researched usability attributes of System Acceptability Framework by Nielsen (Nielsen, 1994) that are described below:

- **Supportability**: System’s ability to provide helpful information to resolve the issues.
- **Reliability**: System’s ability to perform the intended functions over time.
- **Compatibility**: System’s ability to perform the consistent actions based on what people do.
- **Learnability**: User’s ability to learn to use the system.
- **Efficiency**: User’s level of productivity after having learned to use the system.
- **Memorability**: User’s ability to reestablish proficiency of the system after a period of not using it.
- **Errors**: System’s ability to make as few errors as possible or easily recover from them.
- **User Satisfaction**: System’s ability to provide a pleasant user experience.
- **Intention to use**: User’s perceptions about using the system in future.
- **Utility**: System’s relevance to achieve the desired outcomes.
Measuring system acceptability level

Having good usability does not assure that the application is eventually received and used by its targeted population, therefore it was also needed to analyze a number of factors that could influence the decision of dyslexic child about How and When to use the application. To assess the issues affecting dyslexic child’s level of acceptance of technical solution, we applied the well-grounded aspects of Technology Acceptance Model by Davis (Chuttur, 2009) that is defined as:

- **Ease of Use**: Easy to understand and use.
- **Usefulness**: Fitness for purpose.
- **Intention to Use**: Users’ plan to adopt in future.
- **Attitude**: Users’ expression of favor or disfavor.
- **Motivation**: Users’ desire for an action.
- **Appearance**: Users’ preference about visual looks.

Evaluation procedure

All participants were equipped with real application installed on Galaxy SIII tablet for trial purpose in the presence of their teachers who encouraged them to use the application. They were allowed a period of two weeks to explore all features of the application. The teaching staff elicited the responses of dyslexic children through observation while they were interacting with the application. This provided us with considerable evaluation outcomes which helped modify our application. Some questionnaires to participate in the research were also sent by email to the distant remedial education institutions.

Formative evaluation results

The participants used the application to demonstrate their knowledge of English language and mathematics by performing writing activities. The feedback reported here was a monitored formative evaluation collected through observations by remedial education teachers to analyze the learning effectiveness of application. Collectively, they found that evaluation module was able to give a clear picture of to what extent a dyslexic student has learnt a certain concept by measuring and sharing the learning graph of his/her performance for each lesson activity. Almost 92% of participants were able to complete a lesson activity correctly after attempting it four times. The feedback also revealed the efficiency of history module to demonstrate the advancements in dyslexic student’s writing skills by highlighting the most effective lesson plans and taking a closer look at those that are not measuring up. 89% of participants showed fine improvements in their performance score and achieved their targeted learning objectives after five learning sessions. Others required more time due to their carelessness or difficulty to stay focused during learning activities. Furthermore remedial education teachers interpreted that presenting writing concepts through audiovisual elements enhanced the learning outcomes of participants. They mentioned that one of the application areas to be improved was level of interactivity as the learning effectiveness of application had greatly depended on it. Also verbal prompts were required to teach the dyslexic children to interact with the application.

Analysis of system/user performance results

By evaluating the application with a group of dyslexic students under the supervision of their teachers, the questionnaire helped to provide their valuable feedback relevant to design choices of the application in enhancing their handwriting skill. The feedback data was computed through Excel Spreadsheet for the analysis purpose. Figure 6 depicts the statistical analysis of questions included in User Interface section having the overall mean score of 4.37 which indicates that 87% of participants agreed that design choices of application interface are appropriate for learning the handwriting skill. Although the evaluation result data is preliminary, analysis revealed that dyslexic students got familiar with features of application in a short usage period as they were able to operate the application effortlessly. The electronic version of the writing tasks delivered a pleasant experience to students which intrigued them to use the application. They also showed increased preference in learning as well as practicing the writing skill on the application rather than on paper. Both the graphical and textual attributes of application helped dyslexic students to concentrate, avoid distractions and target their attention on the application’s screen. Moreover, they easily recognized and utilized all the features of the application independently.
Analyzing the feedback results of experimental evaluation uncovered the student’s progress relevant to learning or improving the handwriting skill. The statistical analysis of questions belonging to Learning Content section of questionnaire is shown in Figure 7 where overall mean score is 4.34. This depicts that almost 87% of dyslexic students clearly favored the structure of the learning content implemented in the application. They showed significant progress, not only in learning the writing and reading skill but also in phonological skill over a short period of time. The students also showed great interest in completing the writing tasks with stylus pen on application rather than on paper with pencil. The syllabus design directed to conventional educational content and dyslexia friendly text layout helped dyslexic students to recognize, identify, memorize and write the characters in an effortless way. The instructional help provided was supportive and easy to comprehend thus they did not ask for any writing assistance.

The quantitative feedback of evaluation process also highlighted the promising prospects of application regarding assessment and reporting of dyslexic’s learning performance. Figure 8 demonstrates the statistical analysis of questions related to Learning Assessment section with overall mean score of 4.42 which evidently states that 88% of participants are in agreement with the efficient usage of application to provide practical experience. Most of the students showed lower frustration level and were found involved with practicing and completing the writing tasks on application as compared to writing on paper. They were also comfortable with interpreting their achieved scores. The representation of learning graph motivated the students to enhance their results in each next attempt. The revision activities provided by application stimulated the ability of dyslexic
students to remember and recall the information without putting excessive pressure on their memory. In addition, the prominent appreciation of the learning efforts provided by application resulted in an increased sense of self confidence.

Analysis of App Learning Evaluation

![Figure 8. Result analysis-learning evaluation section](image)

**Analysis of system acceptability results**

Evaluation of developed application from dyslexic children within the learning setting helped to reveal the degree of application’s significance in terms of design choices, usage as well as adoption in the future. The feedback data of questionnaire that was carefully developed, based on research survey, measured the compliance of the specific parameters of employed technology acceptance model. The mean score for various criterion of preferred model is shown in Figure 9.

![Figure 9. System acceptability results](image)

Dyslexic students gained much experience in a short period of time and got comfortable moving around the application. This is indicated in the mean score for Ease of Use which is 4.35 clearly showing that 87% of them found it convenient to navigate and distinguish the appropriate icons within the application for desired action. The effectiveness of application in dyslexic’s leaning process is demonstrated in mean score for Usefulness.
which is 4.46 stating that 89% of dyslexic students found the application as a pragmatic assistive tool to enhance their understanding and productivity of handwriting skill. The mean score for Intention to Use is 4.42. This means that using the application boosted the interest and confidence of almost 88% of dyslexic children towards learning and practicing the handwriting skill. For Attitude, the mean score is 4.40 showing that 88% of them are in favor of using the application in view of the fact that its interface design and educational content structure potentially helped dyslexic students to improve the foundational skills of handwriting. A prominent increase was noticed in overall performance scores and involvement of dyslexic students with application over brief period of time which also reflected in mean score for Motivation which is 4.31. This result shows that 86% of the students with dyslexia found the learning activities incorporated in application engaging and appreciative. This reflected in the mean score for Appearance which is 4.34 depicting that 87% of dyslexic students found the application to be attractive and consistent in terms of the implemented graphical and textual attributes.

Conclusion

By utilizing the modern mobile technology features, we have developed an application addressing the diverse needs of dyslexic children with learning differences focused on improving their introductory skills of handwriting. It was designed and evaluated with the collaboration of dyslexic children, their parents and remedial teachers having hands on experience in field of dyslexia, to ease the learning difficulties of dyslexic children as well as to measure progress in their handwriting skill over a brief period of application usage. The preliminary results of application evaluation showed the promising effectiveness of developed application in advancing the handwriting skills related to English language and mathematics as significant progress was reported by WLA in the overall learning performance of dyslexic children which otherwise was hard to observe.

The initial evaluation yielded that both the user interface design and learning content structure of application fulfills the elicited requirements of dyslexic children aged 5 and under. Examining the impact of developed application on children with writing differences highlighted that the usage of devised WLA served its purpose within the research scope by improving learning performance of dyslexic child, motivating them to enhance the learning growth over period of practice and eliminating the need of any writing assistance to large extent. Thus the design of WLA focused on measuring the learning progress of children with learning differences is believed to be a great addition to the existing literature as there still is a lack of reliable and valid performance measurement tool to accomplish this job. Since the application implements a great deal of learning material design based on educational aspects of Pakistan’s special education sector, therefore it expected to prove as an efficient and cost-effective technology-based educational resource solving the issues of educational technology awareness, affordability and accessibility in Pakistan.

Limitation and future work

The application was developed on android platform, following the industry standards and guidelines, to provide a cost-effective technological learning solution for dyslexic children. But to accomplish so, we abandoned a large user population currently operating the mobile technology based on other platforms. Therefore, the educational system developers should focus on expanding the audience by developing the application build on other platforms e.g., iOS and Windows. Also they should consider integrating the application into other mobile devices rather than tablet, after the careful evaluation of specific mobile device.

One of the essential extensions worth of research would be to include the writing assessment of two and three lettered sight words in the application to remediate the dyslexic issues related to spelling concepts. Furthermore we intend to perform the summative evaluation of application over time period of six to twelve months providing substantial evidence relevant to its effectiveness in learning and/or improving handwriting skill. We will accommodate dyslexic students enrolled at other remedial education institutes to get more consistent results. After improving the application based on evaluation results, we plan to disseminate it through the online android market. This will enable android users to acquire the application for individual incentive such as by parents who want to get it for their dyslexic child to use outside the classroom setting or for shared incentive such as by remedial educators and dyslexia educational institutes to use within the mainstream classrooms.
References


Promoting Student Metacognition through the Analysis of Their Own Debates. Is it Better with Text or with Graphics?

Marc Lafuente Martínez1* and Ibis M. Álvarez Valdivia2

1Fundació Jaume Bofill, Barcelona, Spain // 2Educational and Developmental Psychology Department, Universitat Autònoma de Barcelona, Spain // marc.lafuentem@gmail.com // ibismarlene.alvarez@uab.es

Corresponding author

(Submitted July 1, 2015; Revised November 4, 2015; Accepted January 29, 2016)

ABSTRACT

This paper presents a higher education experience aimed at explicitly promoting metacognitive processes in a social and collaborative context. Students carried out a debate on an e-forum, and were later asked to collaboratively analyse their own debates. The control group conducted this analysis using text-based tools; the experimental group analysed it with a graphical tool (“DebateGraph”). We examine the consequences of such experiences in promoting students’ metacognitive processes for argumentative competence, as well as its impact on content knowledge learning. The analysis yields different results depending on the perspective adopted: students’ self-assessment or instructor’s assessment.

Keywords

Metacognition, Argumentation, Graphic representation, Information and communication technologies, Collaborative learning, Higher education

Introduction

The ability to argue is an important competence in many current education systems (Rapanta, Garcia-Milá, & Gilabert, 2013). Argumentation, defined as the valid combination of claims and premises (Plantin, 1996), is promoted under the assumption that it leads to the construction of more meaningful knowledge by means of metacognitive processes. Argumentation, developed either in monological or dialogical forms, is often related to the building of more functional and meaningful knowledge (Venville & Dawson, 2010). It is also associated with fostering the development of metacognitive abilities that lead to the control and appraisal of thinking processes, such as the critical revision of perspectives and the refinement of reasoning (Cross, Taasoobshirazi, Hendricks, & Hickey, 2008).

With the emergence of the socio-cultural approach (Nussbaum, 2008; Vygotsky, 1978), social and dialogical forms of argumentation have gained importance. Argumentative dialogue, often examined in collaborative learning settings, is viewed as a tool to promote the mutual regulation of thought processes among learners themselves. This mutual regulation of thought through language is viewed as a means of reaching further levels of thought that the student would not be able to achieve alone.

Debates are a common activity in secondary and tertiary education, and they are often aimed at promoting dialogical argumentation. While debates can be developed with the goal of pursuing the learning benefits described above, some authors (Tumposky, 2004) claim that debates have serious drawbacks, namely, oversimplifying knowledge, presenting false dichotomies or fostering win/lose scenarios in the classroom.

Although debates are commonly developed using oral and written language, with the advent of Information and Communication Technologies (ICT) new forms of representing arguments are available. In effect, ICT enables the representation of arguments in non-textual formats, supporting graphical representations in the form of schemes, tables or visualisations (Noroozi, Weinberger, Biemans, Mulder, & Chizari, 2012). When students are asked to use graphical features to present or analyse their arguments, they can employ technological programs as cognitive tools to modify the representational context of the task (Ertl, Kropp, & Mandl, 2008). As tools are a key element in human activities (Engeström, 1987), the selection of one over another is likely to influence the outcome. Indeed, the use of graphical tools for supporting argumentation has shown diverse benefits, such as the clarification of arguments, keeping arguments on track, giving a general overview of the argumentation’s structure, or helping find new patterns of evidence (Noroozi et al., 2012).

In this paper, we present a study developed for higher education aimed at explicitly promoting metacognitive processes in a collaborative context. Teams of students collaboratively analysed their own previously held debates. They did these analyses using two different strategies: the control group had to perform the analysis using textual tools; the experimental group had to analyse it with a graphical tool (“DebateGraph,” see...
We examine the consequences of using different knowledge representational tools in promoting students’ key metacognitive processes for argumentative competence. For this end, we take into account both the students’ perspective (considering their opinions and subjective experiences) and the instructor’s perspective (considering her assessment of students’ products). We also examine the cognitive consequences of such an experience in terms of declarative content learning, considering the students’ exam outcomes.

Background

Argumentation and metacognition

Argumentation, conceived as a dialogic form of discussion, can be seen as a social activity that has two goals: first, to support one’s own position by providing evidence and favouring arguments; second, to undermine the opponent’s position by identifying and challenging weaknesses in their argument (Walton, 1998). Achieving these goals requires awareness of one’s own ideas and the other’s position, seeking strengths in your own position and weaknesses in an opponent’s arguments. These processes, which lead to thinking about thoughts themselves (Leitão, 2007), rely upon human metacognition. Metacognition can be defined as the cognition of cognition; it includes at least two main processes: knowledge of cognition and control over cognition (Flavell, 1979). Argumentation depends on and, further, develops the students’ prior metacognitive skills, with argumentation competence requiring three different “metaknowing” components (Rapanta et al., 2013):

- **Metacognitive knowing**: being aware of the appropriate knowledge to support and construct arguments (know-what); metacognitive knowing mainly refers to declarative knowledge (e.g., what concepts can I use to support my stance? What evidence is appropriate to support my argument?).

- **Metastrategic knowing**: knowing suitable strategies, in accordance with the task requirements, to construct arguments (know-how); metastrategic knowing refers to procedural knowledge, and involves understanding and awareness of the task requirements in order to select appropriate strategies (e.g., what procedures can I use to better fulfil the argument task? On what basis can I decide whether an idea is right or wrong?).

- **Epistemic knowing**: being aware of the consequences derived from the cognitive performance in an argumentative task (know-be); this involves knowing about knowledge, both in general and in relation to individuals (e.g., has the argument provoked any knowledge advance in participants? Is argumentation a good setting for solving mathematical problems in teams?).

Argumentation in educational settings relies heavily upon prior metacognitive skills development in students, and instructors. However, argumentation can also be regarded as providing the ideal context to promote such skills in students.

Given that instructors have to be aware of the metacognitive processes involved in argumentation in order to promote them in their students, argumentative tasks have been considered as an excellent method for teacher-training. Argumentative activities are likely to promote Student teachers’ metacognitive abilities. Moreover, with the emergence of ICT, asynchronous forums have proved to be a cost-efficient and generally effective activity for this purpose (Topcu, 2010; Topcu & Ubuz, 2008). Although synchronous discussions have also been used by Student teachers (Chen, Chen, & Tsai, 2009), asynchronous argumentation seems to clearly promote higher metacognitive-knowledge strategies through messages, such as, exemplification, clarification or elaboration (Topcu & Ubuz, 2008).

Metacognition has mostly been examined as an individual process, rather than a social or distributed phenomenon among members of a group (Goos, Galbraith, & Renshaw, 2002). In a study that involves preservice science teachers in a problem-based learning scenario, Siegel (2012) broke group metacognition down into three components, with the group using these components to reduce the distance between their member’s positions:

- **Metasocial awareness**: the group identifies “who” has “what” expertise and resources;
- **Monitoring understanding**: the group manages to identify what they know on a public level; rendering “holes” in the group’s understanding visible is especially important;
- **Monitoring process**: the group manages to set goals and revise them according to the process they have followed.
Group metacognition may have distinctive properties compared with individual metacognition. Group metacognition is under-researched and its effects on teaching and learning are highly underappreciated (Anderson, Nashon, & Thomas, 2009). Therefore, we must take into account the need to develop and study such processes at a group level.

**Argumentation and knowledge representation**

Over the last 15 years, researchers and practitioners have implemented the use of ICT for supporting argumentation in collaborative activities (Noorozi et al., 2012). Some of these studies used technology to represent arguments in different communicational formats (e.g., graphics or text), and their impact on the collaboration-process and students’ learning outcomes was studied. During these activities, students commonly used technology to scaffold their argumentation, engaging with ICT to represent arguments in a graphical format, such as schemes (Schwarz & De Groot, 2007), tables (Suthers & Hundhausen, 2003), or visualisations (Munneke, Van Amelsvoort, & Andriessen, 2003). Most of these studies compared two conditions (i.e., using graphical representation of arguments versus textual representation), and generally yielded positive outcomes for the graphical condition.

For instance, Dwyer, Hogan and Stewart (2012) claimed that students using argument mapping in an e-learning course gained a significantly larger improvement in critical thinking skills than students in the control group, who used textual tools. Argument mapping improved learning by minimising the cognitive load involved in interpreting arguments. Dwyer et al. (2012) affirmed that argument mapping may enhance metacognitive thinking by making the structure of the argument open to appraisal and discussion, and by revealing strengths and weaknesses in the structure. Similarly, Nguyen (2009) asserted that undergraduate psychology students benefited from using visual features (e.g., discourse maps) in their online discussions. Critical thinking and knowledge construction are enhanced when students are allowed to visualise their entire discussion along with the relationship between messages through clear pictures.

Along similar lines, other studies have supported the use of visual technological features. Munneke, Andriessen, Kanselaar and Kirschner (2007) confirmed that students argue more thoroughly, both broadening the debate space and examining in greater depth, when they use a tool to represent arguments in a diagrammatic fashion. However, students seem not to take full advantage of tools proposed to foster argumentation. Students may lack experience using such tools, or interpret them differently to how the researcher or instructor intended (Erkens, Jaspers, Prangsma, & Kanselaar, 2005). Therefore, diagrammatic representations may improve collaborative argumentation, but only when they are used in a co-constructive way (Van Amelsvoort, Andriessen, & Kanselaar, 2007). Janssen, Erkens, Kirschner and Kanselaar (2010) concluded that diagrammatic representations may have a positive impact on outcomes and subsequent elaborated student outputs, such as written essays, but not less so in relation to the online collaboration process itself. Indeed, students reported more negative evaluations of the social process and more neutral technical remarks. This was interpreted by the researchers as an indication that using a graphical tool rendered the argumentation process more complex to students. As a result, students do not perceive the graphical tools as more useful than a textual tool. Janssen et al. (2010) argued that students’ perceptions of a tool’s usefulness may not correspond to its objective efficacy or efficiency.

Some studies yielded unfavourable results for visual representations. For instance, Van Drie, Van Boxtel, Jaspers and Kanselaar (2005) attempted to promote argumentation in the subject of History by using representational guidance, such as diagrams and matrices. Contrary to expectation, they found that students actually engaged in increased historical reasoning during subsequent discussions, if they had not used representational guidance. Namdar and Shen (2013) presented a study where students could use multiple representations in a virtual environment (textual and visual representations through concept maps, wikis and events entries), in order to engage in collaborative argumentation in physics. Although students used all the available representational modes to make sound arguments, text entries were most frequently employed by the student-teams. According to Namdar and Shen (2013), although students are perfectly able to represent their knowledge in the form of concept maps or pictorial representations, they may still feel the need to convey understanding to their peers through verbal explanations. Although personal preferences in the processing of information might be involved (i.e., visual versus linguistic), some studies also suggest gender differences: given the same conditions, female students prefer to build and use textual digital artefacts rather than visual ones (Ding & Harskamp, 2006).
Methods

Design

This research used a case study approach. The case study is an appropriate method for researchers who want to attain a perceptive understanding of an instructional context, by seeking answers to descriptive and explicative questions (Yin, 2003). Specifically, this study used a quasi-experimental design, with a multi-method approach to the data analysis. Qualitative and quantitative data may be combined, used independently, concurrently or sequentially, embedded into each other, or used as a foundation to develop the other (Creswell & Clark, 2007). It allows the researcher to view the study phenomenon from different perspectives, and therefore be able to use one or various methods in order to address specific questions.

Participants

The study took place on a Developmental Psychology course in a postgraduate programme (a Master’s degree for Second Language Education Teachers) in a public university in Barcelona, Spain. There were 56 students (43 female, 13 male, $M = 25.6$ years, $SD = 5.9$; range = 21-43). A large proportion of the students (49, 88%) had a Philology Bachelor’s degree (33 Catalan Philology, 16 Spanish Philology). The rest had a Bachelor’s degree in Journalism or Audiovisual Communication. Most of the participants (72%) stated they did not have any experience in learning through technology.

Procedure, tasks, and instruments

In this study, we examined three different activities that occurred as part of the same unit: “Development of thinking in adolescence.” The first two activities were developed in collaborative teams of five or six members ($n = 10$); these teams were organised by the students themselves at the beginning of the course. The third activity was an individual exam.

The first activity was a forum-based debate developed within each group. On the Moodle platform, the groups had their own work space within the debate area to carry out critical discussions. They received the same instructions from the teacher, and were assigned the same amount of time to develop the debate. The students were presented with a video-clip that showed possible ethnic conflicts in the context of relationships between adolescents and adults, which stemmed from adolescent thought characteristics. The students were asked to adopt the role of a secondary education teacher, and discuss the viewpoint that they should have as educators: “Must the teacher promote any specific ideology in the adolescent? And a religious stance? To what extent are we entitled to promote an ideological or political change in the student?”

With respect to the participation norms, each student had to submit at least two messages: one to present new ideas, and another one to react to peers’ contributions. There was no limitation to the number of contributions allowed. The instructor recommended ideas and arguments were grounded in unit content (Development of thinking in adolescence). The debate was open for two full weeks.

In order to assess the virtual debate’s quality, the instructor took into account the quality of the ideas elaborated by the students, the argumentation and support of the ideas, and the accomplishment of the debate instructions. In accordance with those criteria, every debate was evaluated and scored using a three-point scale, where A was excellent, B was good, and C was acceptable/passing. Students’ debate quality was used in their assignment to the experimental or control condition, alongside their gender and experience in learning through technology. All these variables were well-balanced.

The second activity took place two months after the end of the virtual debate, in a two hour face-to-face session. The aim of the second activity was to analyse the previous debate. Half of the groups ($n = 5$) were assigned to the experimental condition, where they used the tool DebateGraph. The experimental groups had the support of another instructor who, in the first 45 minutes of the session, trained the students in the use of the tool. The other half of the groups ($n = 5$) were assigned to the control condition, where they did not receive any specific instructions regarding tools. In order to present their analysis, all the control groups used the text editor Microsoft Word.
During this activity, students in both conditions used a printed record of their messages that was handed out by their instructor. The messages were ordered chronologically and organised by thread. The students received the instruction to identify and present the main ideas or standpoints that emerged in their forum. They then had to identify and present the main counter-ideas or counter-standpoints that emerged; and finally, to identify and present the main arguments that arose to support all the key ideas and counter-ideas.

The third activity was an individual exam, where the students were presented with an open question that assessed the conceptual knowledge discussed in the debate: “Explain the main features of adolescent thinking.” The answers were graded by the instructor on a 0 to 10 scale point, according to the theoretical framework of the course, the epistemological theory of Jean Piaget. Below, in Table 1, we provide the criteria used to assess the answers.

Table 1. Criteria used to assess the students’ exam answers

<table>
<thead>
<tr>
<th>Grade</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No answer</td>
</tr>
<tr>
<td>1-4</td>
<td>Incorrect answer: no contents addressed or anecdotic / no theoretical or conceptual framework provided / vague or ambiguous answer / important conceptual mistakes in the answer</td>
</tr>
<tr>
<td>5-6</td>
<td>Acceptable answer: the answer presents most of the features but lacks the conceptual framework / the student gives opinions rather than documented topics / the content is not correctly explained</td>
</tr>
<tr>
<td>7-8</td>
<td>Good answer: the answer correctly presents most of the cognitive developmental features but lacks some of them</td>
</tr>
<tr>
<td>9-10</td>
<td>Remarkable answer: the answer covers all the cognitive developmental features and is developed within an appropriate theoretical framework</td>
</tr>
</tbody>
</table>

To explore the data, we firstly examined the students’ appraisal of their own experience through a questionnaire. This questionnaire addressed different metacognitive processes involved in argumentative competence. Second, for a more detailed approach, we analysed their debate’s analysis products, which also focused on different metacognitive processes. Third, we evaluated the repercussions of the whole experience in terms of specific content learning, by assessing their answers to the final exam question on the unit contents.

Table 2. Questionnaire to assess the students’ perspectives on metacognition

<table>
<thead>
<tr>
<th>Items</th>
<th>Metacognitive knowing (know-what)</th>
<th>Metastrategic knowing (know-how)</th>
<th>Epistemic knowing (know-be)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>The analysis has helped me identify and present the main ideas and counter-ideas of our debate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>The analysis has helped me identify and present the main arguments of our debate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>The analysis has helped me organise and synthesise the content of our debate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>The analysis has helped me acknowledge the main strengths of our debate (e.g. interesting ideas, contrast of relevant ideas, solid arguments, etc.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>The analysis has helped me acknowledge the main weaknesses of our debate (e.g. poor ideas, absence of counter-ideas, weak or non-existent arguments, etc.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>I think that through this activity I have been able to develop skills for debate analysis that perhaps I did not previously have</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>On a scale from 0 to 10, my general satisfaction with the debate analysis experience is…</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. The alpha coefficients for internal consistency averaged .85.

With the purpose of assessing the experience from the students’ perspective, a brief questionnaire was constructed. As we showed in Table 3, the survey covered the different metaknowing strands involved in the student’s debate analysis, according to the conceptual framework presented in the literature review (Rapanta et al., 2013). The questionnaire consisted of 7 Likert-type items with a 4-point response scale; from strongly agree to strongly disagree. In addition, an open question allowed the students to express any comments they judged relevant. The questionnaire is completed online once they had finished their debate analysis. The questionnaire was voluntary and anonymous. 43 students complete the questionnaire (77% of participants), 23 from the experimental group and 20 from the control group. Table 2 shows the metacognitive strands assessed through the questionnaire.

To fulfil the second stage of analysis, we created a rubric, allowing us to evaluate the student’s debate analysis products. The rubric was the product of an inductive-deductive design procedure (Moskal & Leydens, 2000). Firstly, we defined the operational criteria for every category, based on theoretical principles derived from the
literature review. Then, the rubric was revised by the researchers in an iterative process that involved modifying categories. Discrepancies between the researchers were resolved via consensus. The final aspect of examining the analysis involved a cross-case analysis (Miles & Huberman, 1994). The rubric was used by the instructor to assess the debate’s analysis of both the experimental and control groups. That rubric aligns with the theoretical framework adopted in this study, and was used to assess the quality of the students’ analyses, based on the metacognitive, metastrategic and epistemic strands found in their products (see Table 3).

Finally, to complete the third step of the analysis, in the final exam, the instructor included a question relating to the contents of the previous debate. The question assessed the conceptual knowledge constructed by the students as a result of their participation in all prior activities of the unit. The results of the exam question were then compared with the results of the prior steps in analysis.

<table>
<thead>
<tr>
<th>Table 3. Rubric to assess the students’ analysis product</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clarity of ideas, counter-ideas and arguments</strong></td>
</tr>
<tr>
<td>D (1)</td>
</tr>
<tr>
<td>Most of the ideas, counter-ideas and arguments presented in the analysis are unclear or difficult to understand.</td>
</tr>
<tr>
<td><strong>Relation between ideas, counter-ideas and arguments</strong></td>
</tr>
<tr>
<td>Most of the ideas and counter-ideas are not clearly related or opposed to each other in the analysis.</td>
</tr>
<tr>
<td><strong>Quality and appropriateness of ideas, counter-ideas and arguments</strong></td>
</tr>
<tr>
<td>Ideas, counter-ideas and arguments presented in the analysis generally do not address or answer the debate questions.</td>
</tr>
<tr>
<td><strong>Organisation and synthesis of contents</strong></td>
</tr>
<tr>
<td>The product is badly organised and it does not appropriately synthesise the debate.</td>
</tr>
<tr>
<td><strong>Understanding and meeting the analysis requirements</strong></td>
</tr>
<tr>
<td>The students clearly do not understand and meet the requirements of the analysis.</td>
</tr>
<tr>
<td><strong>Understanding and deepening of contents</strong></td>
</tr>
<tr>
<td>The students clearly show incomprehension of the contents, and a superficial conception of the analysis topics.</td>
</tr>
</tbody>
</table>
Results

Our aim was to study the metacognitive consequences of using textual or graphic tools by students when analysing their own debates. We examined whether there were differences between the experimental and control groups, regarding the metacognitive strands involved in the students’ analysis. Below we present the answer to this initial question, firstly from the students’ perspective, taking into account their responses to the questionnaire, then from the instructor’s viewpoint, considering her assessment of the students’ analysis products. We then consider the cognitive consequences of such an experience, analysing whether there are differences between the experimental and control groups in terms of declarative knowledge content learning. To inform this, we consider the instructor’s assessment of the students’ exam answers.

We observe significant differences in the appraisal of the learning experience by the students. All the item scores show a statistically significant difference, with the exception of the item “awareness of development of new analytical skills”. Table 4 summarises the ANOVA conducted to compare both groups.

Table 4. Summary of ANOVA for students’ questionnaires

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Experimental group</th>
<th>Control group</th>
<th>F(1,41)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (n = 23)</td>
<td>M (n = 20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metacognitive knowing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identification of ideas and counter-ideas</td>
<td>2.85 .67</td>
<td>3.34 .64</td>
<td>6.12</td>
<td>.018</td>
</tr>
<tr>
<td>Identification of supporting evidence and arguments</td>
<td>2.8 .62</td>
<td>3.34 .64</td>
<td>8.01</td>
<td>.007</td>
</tr>
<tr>
<td>Support to organising and synthesising contents</td>
<td>2.8 .83</td>
<td>3.47 .67</td>
<td>8.79</td>
<td>.005</td>
</tr>
<tr>
<td>Metastrategic knowing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identification of strengths</td>
<td>2.85 .67</td>
<td>3.57 .59</td>
<td>13.84</td>
<td>.001</td>
</tr>
<tr>
<td>Identification of weaknesses</td>
<td>2.5 .76</td>
<td>3.74 .45</td>
<td>43.64</td>
<td>.000</td>
</tr>
<tr>
<td>Epistemic knowing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awareness of development of new analytical skills</td>
<td>2.35 .93</td>
<td>2.7 .93</td>
<td>1.48</td>
<td>.231</td>
</tr>
<tr>
<td>Satisfaction with the activity</td>
<td>6.35 1.79</td>
<td>7.83 1.44</td>
<td>9.03</td>
<td>.005</td>
</tr>
</tbody>
</table>

Analysis of the debate’s analysis products, as evaluated by the instructor, did not yield any significant differences. As we show in Table 5, when a t-test analysis is conducted, none of the rubric’s criteria show a statistically significant difference. Hence, the tool used to perform debate analysis does not seem to have a significant impact on the students’ product quality. Further, the analysis conducted does not show any difference in any of the metacognitive strands, regarding the analysis products of both groups.

Finally, ANOVA results do not show statistically significant difference between the groups in terms of their exam outcomes. Students’ answers, as assessed by the instructor, in the experimental group ($M = 6.33, SD = 3.26$) and the control group ($M = 7.69, SD = 2.22$), are not statistically significant $F(1,54) = 3.337, p = 0.73$. Hence, the students’ answers to the exam question do not show any difference in terms of acquired declarative knowledge related to the unit’s.

Table 5. t-test: Instructor’s assessment of the students’ debate analysis

<table>
<thead>
<tr>
<th>Variable</th>
<th>Experimental</th>
<th>Control</th>
<th>t</th>
<th>p</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (n = 23)</td>
<td>M (n = 20)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ideas identified</td>
<td>4.0 2.23</td>
<td>6.0 3.0</td>
<td>1.20</td>
<td>.266</td>
<td>[-1.86, 5.86]</td>
</tr>
<tr>
<td>Counter-ideas identified</td>
<td>2.2 1.3</td>
<td>2.4 .89</td>
<td>.28</td>
<td>.784</td>
<td>[-1.43, 1.83]</td>
</tr>
<tr>
<td>Arguments identified</td>
<td>6.8 2.49</td>
<td>7.6 2.1</td>
<td>.55</td>
<td>.596</td>
<td>[-2.54, 4.14]</td>
</tr>
<tr>
<td>Metacognitive knowing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clarity of ideas and counter-ideas</td>
<td>3.8 .45</td>
<td>3 1.0</td>
<td>-</td>
<td>.141</td>
<td>[-1.93, .33]</td>
</tr>
<tr>
<td>Relation between ideas and counter-ideas</td>
<td>3.8 .45</td>
<td>3 1.09</td>
<td>-</td>
<td>.290</td>
<td>[-1.82, .62]</td>
</tr>
<tr>
<td>Quality and appropriateness of ideas and</td>
<td>3.6 .55</td>
<td>3.2 .83</td>
<td>-.89</td>
<td>.397</td>
<td>[-1.43, .63]</td>
</tr>
</tbody>
</table>
counter-ideas
Organisation and synthesis of contents  3.8 .44  3.6 .54 -.63 .545 [-.93, .53]

**Metastrategic knowing**

Understanding and fulfilment of the analysis requirements  3.8 .45  3.6 .55 -.63 .545 [-.93, .53]

**Epistemic knowing**

Understanding and deepening of the contents  2.6 .54  2.8 .84 .45 .667 [-.83, 1.23]

*Note. CI = Confidence Interval.*

**Discussion**

Our study yielded some interesting results on how graphical and text-based tools promote the metacognitive skills involved in argumentation. With regard to the accounts of students themselves, gauged by our questionnaire, the students using the graphical tool reported a significantly worse appraisal of metacognitive abilities fostered in their debate analysis. Indeed, according to their subjective experience, the students using text-based tools reported high scores in all three declarative metacognitive (know-what) components included in our questionnaire: identification of ideas and counter-ideas; identification of supporting evidence and arguments; and supporting the organisation and synthesis of the debate. Therefore, we may state that the students felt that their debate content is better appraised and acknowledged when using text-based tools to analyse it. Similarly, the students using text-based tools showed significantly higher scores related to metastrategic knowing (Rapanta et al., 2013). Know-how metacognition encompasses showing awareness of the debate task and assessing whether the procedures used for constructing arguments are valid or not. This process, if it is developed in a distributed fashion within the group, may also be related to an important element of group metacognition, namely, monitoring understanding (Siegel, 2012). Metastrategic knowing was appraised in our questionnaire by two different items: identification of strengths, and identification of weaknesses. Hence, we can affirm that students believed that using the graphical tool is not as useful as text-based tools, when it comes to analysing whether their argumentation is based on valid strategies. As for the epistemic (know-be) component, the students in both groups showed an equally moderate perception of having developed new skills for analysis after the activity. However, when they assessed their overall satisfaction with the analysis activity, the students using the text-based tools show significantly higher scores again.

In sum, although the students judged that representing their previous debates in a graphical form may fairly support their analysis (most of their average scores are above the mid-point), text-based tools are judged to be more effective for promoting the metacognitive processes involved in argumentative competence (Rapanta et al., 2013). Indeed, according to the subjective experience of the students, text-based tools are more useful in terms of developing new analytical skills, the experimental and the control group converged in giving an equally moderate score. These results reflected the conclusions of previous research (Janssen et al., 2010), where students reported a rather negative appraisal of the collaboration process and the role played by a graphical digital tool. The fact that the students were not experts in such a program, and that they might have idiosyncratically interpreted the goal of using the tool, for example, not as a means to foster metacognitive processes but simply as a means to draw their arguments, could also explain these results (Erkens et al., 2005). The limited time invested in the classroom for mastering a new tool, or the students’ effort put into abstracting and making the structure of their debate explicit (Dwyer et al., 2012), could have decisively tainted their experience. After all, students may have much more experience reading and writing text-based arguments, compared to graphic-based arguments. Prior research has identified that students feel generally more comfortable using text for representing and communicating their knowledge to peers (Namdar & Shen, 2013). In addition, female students may have a tendency to construct and use text digital
artefacts rather than graphical ones (Ding & Harskamp, 2006). Given that our class group was composed of 42 female students and 13 male students, students’ gender could also have played a role in their subjective appraisal of the learning experience. This explanation is supported by the fact that, although the instructor in the control group did not give any specific instructions regarding the written tool for their analysis, they all used Microsoft Word.

Regarding the quality of the students’ analyses, as assessed by our rubric, we can assert that although the experimental group was slightly above the control group in most of the strands considered, the differences are not statistically significant. Hence, their products reflected the same overall level of achievement in all the metacognitive strands assessed (Rapanta et al., 2013). We may speculate that a bigger sample of groups could have led to statistically meaningful differences; however, in our sample those differences are not significant. Therefore, we can affirm that using a graphical tool instead of a text-based tool did not yield any significant differences regarding the clarity of their ideas, counter-ideas, and arguments expressed in their analysis. Likewise, the appropriateness and quality of those ideas, counter-ideas, and arguments were at the same level for both groups. Both groups also showed equal achievement in the relations expressed between debate elements (relations between ideas and counter-ideas, and between arguments and ideas). Finally, when assessing the contents’ organisation and synthesis, both groups showed an equally excellent level.

Given that we find no significant differences between the experimental and the control group, we cannot corroborate the advantages identified by previous research in the use of graphical formats for supporting argumentation (Noroozi et al., 2012). The benefits found by others, such as enhancing metacognition by making the structure of the argument more open to appraisal, and by revealing its strengths and weaknesses (Dwyer et al., 2012), or broadening and deepening the debate scope (Munneke et al., 2007), cannot be supported by our study. However, we must consider that the approach adopted in this study is different to that of other studies: while previous studies usually propose using a graphic tool for supporting consequent (or simultaneous) argumentation, we used the graphical tool as a metacognitive tool to analyse previous argumentation. Using a tool as a means of fostering evaluation of one’s previous argumentation might be more demanding than simply considering it as a support for argument planning or execution. After all, evaluation of one’s cognition is clearly a different process than planning and monitoring (Meijer, Veenman, & Van Hout-Wolters, 2006), as it requires demanding operations to revise one’s behaviours and thoughts, in relation to specific goals. Therefore, exploiting the tool to evaluate personal and group understanding (Siegel, 2012), and measuring those advantages, might be more difficult than using it for planning or execution purposes. We might speculate that the advantages of using external representations in graphic formats may be more easily fostered and measured when the tool is used to promote subsequent argumentation, but more difficult to determine when the tool is used to evaluate previous argumentation.

It is noteworthy that although the experimental group showed a more negative subjective appraisal of their experience, when we assessed their products, their results were at least as positive as the control group. In our view, this reinforces previous conclusions that students’ perceptions of tool advantage may not correspond to objective effectiveness (Janssen et al., 2010). Students’ perceptions may be driven by their struggle to master a new digital tool (72% of students did not have any experience in learning through technology); therefore, drawing their attention to the tool’s limitations, rather than its affordances. Likewise, the positive experience reported by the control group may have more to do with working with a familiar and comfortable form of representation (i.e., text), than with its objective consequences (Namdar & Shen, 2013).

Finally, regarding the student’s acquisition of declarative knowledge, it must be asserted that using either tool did not make any difference in learning ideas, concepts, and facts from the unit. Hence, using text or graphic tools did not have any impact, neither on the “higher” metacognitive processes developed while analysing their own debates, nor on the “lower” cognitive processes carried out while learning the ideas and concepts used in those debates.

Conclusions

In our study, students did not take advantage of using a graphical tool to enhance their learning while analysing their own previous debates. Students in the experimental group neither improved their metacognitive nor their cognitive processes in comparison with the text-based tool group. Students using text-based tools reported a more enriching learning experience, but, according to our results, this may be caused by tool familiarity and prior skills of students, rather than real learning outcomes. Our study shows that while graphic tools may enhance learning when used to prepare or execute students’ argumentation, it is not clear that it is beneficial for
analysing previous debates. Indeed, analysis of arguments is different from planning, and requires different skills.

The present study reminds us that while the tool (either digital or analogue) may be an important element of the educational activity, many other variables in the educational setting may interact with each other, and eventually play a role in learning processes. In our study, students’ learning styles or preferences (e.g., visual or linguistic), or even their gender, may have played a role in the results. However, we must accept, as a clear limitation of the present research, that those variables were not controlled. Likewise, we must admit that small-scale studies, such as ours, do not provide a representative sample for generalising results. We urge further research to be conducted, where these limitations can be overcome.

Innovative educational experiences may be rewarding for all stakeholders, as long as they are adequately planned and implemented. A sufficient amount of time and resources must be allocated for instructors and students to learn new tools and scenarios. Likewise, innovative experiences have to be adequately assessed. Our study shows that both students’ and instructors’ perspectives must be taken into account to fully evaluate the soundness of innovation in education.

References


Advancing Adventure Education Using Digital Motion-Sensing Games

Ju-Ling Shih¹ and Yu-Jen Hsu²

¹National University of Tainan, Taiwan // ²National University of Chiayi, Taiwan // juling@mail.nutn.edu.tw // hsuyjen@mail.nctu.edu.tw

*Corresponding author

(Submitted July 3, 2015; Revised February 19, 2016; Accepted March 18, 2016)

ABSTRACT
This study used the Xbox Kinect and Unity 3D game engine to develop two motion-sensing games in which the participants, in simulated scenarios, could experience activities that are unattainable in real life, become immersed in collaborative activities, and explore the value of adventure education. Adventure Education involves courses that emphasize first-hand learning experience in which participants cooperate in situated environments to achieve various life-learning aims. Using qualitative data to supplement questionnaires, system records and observations as well as interviews, results showed that digital games can substantially improve the learning effectiveness of adventure education themes just as traditional physical adventure education activities. Participants can advance their extrinsic behaviors, such as leadership and cooperation, and intrinsic motivations, such as reflection and confidence, through the games. With appropriate guidance and reflection, participants can develop their conceptual and social self. Motion-sensing games are convenient and effective tools for advancing physical adventure education activities.

Keywords
Adventure education, Digital game-based learning, Motion-sensing game

Introduction

Adventure education involves various adventurous and challenging types of group activity, and was established by the Outward Bound School in the United States and Europe through the field activities of boy-scout courses. In 1970, Jerry Pieh and Gary Baker used various types of simple portable tools in their Project Adventure course and integrated outdoor activities into general courses that enable students to learn teamwork, communication, leadership, and introspection (Prouty, 1990). Adventure education emphasizes learning through experience and cultivates the self-confidence, interpersonal interaction ability, and self-development of students. The activities are often designed with rules that limit the communication channels of participants, requiring them to observe their surroundings and empathize with the people around them. However, the effects of the activities decrease when the situation becomes disordered. The facilitator must therefore be proficiently equipped with knowledge and skills to guide participants through the activities and reflect on the problem-solving process to communicate the themes of the activities. The observation duty of the facilitator is critical because one teacher can lead up to 30 participants at a time. Key actions, communication processes, and events must be recorded in a short time.

Some adventure education activities are difficult to practice because of poor weather conditions, outsized equipment, or dangerous scenarios. Therefore, digital games were developed in this study to create simulated scenarios through which players can experience activities that are seldom performed.

Digital games have been used for learning in various fields. As technology and creativity have advanced, game types and interface designs have changed substantially. Games can be played using keyboards, mouse, and remote motion sensors. Microsoft released a kinesthetic game console, Xbox Kinect, which revolutionized the control rules of traditional games. Motion-sensing technology can detect and position the movement of players so that they can operate the game system directly without physical contact. Instead of a handheld or stampede controller, players use voice commands or physical gestures. A built-in camera is used to capture the movements of a player in a 3D space, thus enabling intuitive interaction with the characters and objects in the game. Virtual reality provides players with a sense of novelty, increasing their motivation to learn.

This study involved using an Xbox Kinect and Unity 3D game engine to develop two multiplayer adventure education games. The games addressed cognitive, metacognitive, motivational, and behavioral aspects of learning. The games were developed based on adventure education activities to enable participants to enhance their interpersonal relationships through group interaction. The digital games emphasized learning processes that involved user-initiated control, enabling users to construct situations that fit their intrinsic cognitions, thus improving their learning effectiveness. The participants enhanced their learning and problem-solving capabilities in the process (Perry, 1998).
Through the postgame reflections, the participants were guided in perceiving their roles in the group, understanding the mechanism of leadership, and collaboratively solving problems. The collaborative gaming processes were recorded using computers, and researchers conducted observations. Questionnaires were used to cross-analyze the achievements of the participants based on adventure education themes, which was the primary goal of this study.

**Literature review**

**Adventure education**

Adventure education is an experiential learning process in which people construct knowledge through direct experience, practicing skills, and strengthening values. Kolb (1984) defined learning as the process of transferring experience to knowledge. This theory presents a cyclical model of learning consisting of four stages, namely concrete experience, reflective observation, abstract conceptualization, and active experimentation, among which reflective observation can be considered the most crucial. Kolb and Kolb (2005) stated that learning is a continuous process that occurs through interaction with the environment. Such a process differs from formal education from which people interact with the environment, objects, and people. In a way, using virtual reality games to simulate the physical world which the players cannot attend to in their real life is considered as extended application of experiential learning. Taff and Ahmad (2007) stated that people learn interaction strategies in groups and apply them in their daily lives. Through group interaction, individuals learn to create and reflect. They learn to adapt to society by solving complex problems in life. It is more enjoyable and can increase learning motivation.

Adventure education employs a thematic design model by using physical learning materials in predesigned learning contexts to guide participants toward discovering problems and identifying solutions. In the problem-solving process, participants are encouraged to conduct interpersonal communication to develop confidence through team effort and collective responses to frustration. A group is a microcosm of a society. Experiential learning emphasizes four aspects: practice, methods, teaching, and processes (Doering, Miller, & Veletsianos, 2008). Through activities, simulated situations enhance participants’ understanding of abstract concepts. In a step-by-step procedure, the focus of course content progresses from personal growth to cooperative problem solving. Through continual group interaction, the class atmosphere improves. The cooperation process enables participants to attain a sense of achievement and understand the advantages of group cooperation, and increases their skills, confidence, and knowledge in taking actions.

Human (2006) considered adventure-based activity a crucial aspect of outdoor education. Tsai and Liao (2008) emphasized that outdoor experiential learning involves a series of group activities that guide participants toward discovering individual differences, promote observational learning, increase interpersonal trust, and enhance group development. The adventure education activities can be categorized into three levels depending on the content. High-risk activities include mountain climbing, rock climbing, rafting, wilderness survival, and solitary camping. Mid-risk activities include low and high rope courses. Low-risk activities include group development activities that guide participants in becoming acquainted with each other, communicating, cooperating, solving problems, and trusting each other. Several of the aforementioned activities involve unstable variations in weather, inaccessible locations, and predefined mental limitations. Thus, simulated environments and scenarios can be used to create more experiential learning opportunities for participants.

**Social interaction in situations**

Brown, Collins, and Duguid (1989) introduced situated learning in their paper “Situated Cognition and the Culture of Learning.” Situated learning was developed based on constructivism, suggesting that learners should learn knowledge and skills in real-life situations by using social interaction to create reasonable and meaningful interpretations on which knowledge can be based. However, creating real-life learning contexts by using the limited classroom space and resources available to formal education is extremely difficult. In addition to situational authenticity, learning content that encourages learning transfer is essential for enabling students to learn problem-solving strategies and for enhancing students’ capabilities.

In social learning theory, three elements, people, the environment, and behavior, are considered interrelated. These elements coexist and cannot be arranged in an order or a hierarchy; this reciprocal relationship is called interaction
Cooperative learning is a learning method that involves dividing students into groups and allowing them to work together to enhance learning. The method can be used to teach specific content, strengthen cognitive processes, and enhance learning achievements (Johnson & Johnson, 1990). Studies have determined that cooperative learning consists of an external award and goal structure that can stimulate inner learning motivation. Slavin (1990) indicated that rewards and goals are two intrinsic factors that promote group members to collaborate to achieve goals. Only when group goals are achieved can personal goals be attained. Conversely, for individuals to achieve their personal goals, group members must collaborate closely. Such an interdependent relationship forms the structure of interpersonal interaction. Knowledge is the result of social interaction and must be internalized to become integrated into personal schema. Therefore, discussion and reflection on aspects of the social interaction process should be emphasized.

Collaborative learning can engender such a learning situation and facilitate the construction of knowledge in groups. Collaborative learning occurs when group members possess collaboration skills that can be used effectively. As a positive interdependent relationship is established, group members internalize the skills through repeated practice. Self-regulation is another function of social learning theory.

Digital game-based learning

Digital games refer to all games designed using digital technology, software, or programming (Nanjappa, 2001). Digital games have entertainment functions and can benefit learning (Hwang, Wu, & Chen, 2012). Lee and Chen (2009) mentioned that digital games are a tool that participants can use to develop problem-solving abilities. Hämäläinen (2008) and Mann et al. (2002) have asserted that digital games facilitate education because they enhance and stimulate learning motivation, eliciting active participation in learning activities. Therefore, digital games can be integrated into the classroom to provide participants with opportunities to develop their critical and creative-thinking ability. Thus, using cooperative problem solving as the goal of a digital game is a meaningful endeavor (Paraskeva, Mysirlaki, & Papagianni, 2010).

Digital game-based learning entails using digital games to transform learning experiences and feedback into learning content. Petranek et al. (1992) stated that the feedback generated by a game increases the problem-solving abilities of participants. Players can solve problems in the game to obtain feedback (Gredler, 1996; Salen & Zimmerman, 2004) and can reflect on failures resulting from their mistakes. Therefore, in the games, participants learn through trial and error. Kraiger, Ford and Salas (1993) indicated that digital games can be used in the teaching process to provide students with deep impressions. In the game environment, students can continually attempt various problem-solving solutions without incurring injuries from mistaken decisions. Among all technology-mediated learning activities, digital game-based learning has been proved to be an effective method for increasing the learning motivation of students (Tüzün, Yilmaz-Soylu, Karakus, & Kizilkaya, 2009).

Motion-sensing games developed in this study involves incorporating adventure education issues into digital games, enabling learners to immerse themselves in the situation, relieve their subconscious, experience and feel themselves in the gaming process, reflect on their inner characteristics, fully express their emotions and ideas, and take constructive action to resolve problems and mature (Hsu, Lin, & Shih, 2013). Digital games have functions such as metaphorical teaching, interpersonal communication, self-involvement, game immersion, and affective expression. An advantage of digital game is its system records, which enable facilitators to analyze group dynamics effectively. The digital game developed in this study enables participants to discover and develop their extrinsic behaviors and intrinsic motivations. The reflections of the participants indicated that effective gaming process enhanced their communication and problem-solving skills and facilitated clarifying group goals. Because the participants were willing to share and support each other, they exhibited a high action rate and stable emotional conditions.

The game is also an educational guidance model that is related to situations in daily life. The game design was based on the rules of face-to-face adventure education activities. The content is embedded in the game to simulate real-life situations. Participants can generalize, learn about themselves, and discover their potential to stimulate personal growth. Consequently, theories such as positive reinforcement, observational learning, and behavioral shaping were applied in designing the game in this study.
In the game, learners interact with each other in a positive environment and are immersed in the social process. The system generates positive feedback and virtual rewards to encourage positive behaviors. When personal behaviors result in positive collaboration, the behaviors are carried through to other conditions and learning transfer occurs immediately. This signifies that the problem-solving process is effective and that the participants should address similar situations by using this process. However, because situations have variations, distinctions must be made to enable people to address complex human interaction and relationships. Repeated game play generates various social interaction processes, enabling each player to generate new behaviors progressively to achieve new goals. Learning transfer is speculated to occur through this process.

Game design

Two digital adventure education games, Corporate Maze and Group Balance designed and developed in this research, were based on the physical adventure education activities that are normally conducted to teach participants how to collaborate, solve problems, and demonstrate leadership. Both games require participants to stop using one of their senses during physical interaction. Converting these games into a digital form does not affect their effectiveness. The games are designed to provide learning situations that enable learners to experience cognitive reconstruction and engage in problem solving. Learners reproduce the behaviors in the game to represent the concept of games.

Because the game involved pathfinding, which required trial and error, the system needed to detect only a few body actions to prompt avatar movement. FAAST (Figure 1) was used to differentiate between the positions of the players and the angles of their ankles and translate them into corresponding keyboard events.

OpenNI was used as middleware to connect Kinect and Unity3D. OpenNI (Figure 2) defines the API and provides a cross-platform structure; it is a standard interface used to process gesture recognition and tracking. It also supports multiplayer sensors, which were required in this game to enable collaboration, self-exploration, and social interaction.

The map of Corporate Maze game is a 9 × 5 grid scattered with landmines. Only one safe route passes through the maze. Everyone in the team must take turns to participate in the rescue. Team members cannot remind a player of the accurate route; the player must memorize it. When the player steps on a landmine, the player must return to
the starting point by following the route by which he or she arrived at the landmine or the score is deducted (Figure 3). When the score is zero, the game is over.

![Figure 3. Landmine exploration in corporate maze game](image)

Corporate Maze is played by one group member at a time while other members offer assistance. Players are required to pass through the maze to reach the other side of the board. Whenever they step on the mine, they must take the same route to return to the starting point. Therefore, all members must memorize the correct routes to help each other pass through the maze. The game can be won only when the group members use their collective memory, accumulate past experiences, share information, and communicate.

The setting of the digital game was an open field to release some of the tension of the game (Figure 3 and 4). The exploration effects that occur when players make mistakes are alarming.

![Figure 4. Playing corporate maze game](image)

The goal of Corporate Maze is to develop trust among group members, who must collaboratively make decisions and confront challenges. Through collaborative learning, players can immerse themselves in contextual learning, discover research value through trial and error, and think unconventionally to challenge traditional learning difficulties. At the end of the game, players answer a few simple questions about the game process, such as the route memory strategies, feelings on returning via the same route, concerns regarding the location of the mines and possible solutions, as well as practical applications of the game to life. The players are asked to describe their feelings regarding the unknown and their attempts and to indicate whether they experienced hesitation about breaking the rules and continuing the trip. After the reflection section, the game provides the following feedback.

*Corporate Maze simulates life progression and reminds us that consolidation at every step of life is essential when confronting change. As we pursue success, we must remember the experience of failure. The accumulation of past experiences can improve the future. In the process, we searched for mines, confronted dilemmas, and tolerated exploration; however, with the assistance of peers, communication of solutions, and sharing of experiences, we reached the goal together. When we perceiving a dead end, there is always a way, outside of range, that requires brevity and creativity to find.*

Group Balance is an outdoor high-altitude adventure activity involving a board 45 meters in width hanging at an 85-meter height, with two strings that can pull the board high to reach the goal (Figure 5 and 6). Three group members participate in each round of the game. One member controls the balance of the board and the other two members push the board on the left and right sides, ideally at the same time and with the same strength, to raise the board. The game restarts when the board is dropped. The game requires the members to build trust, observe, and support each other when necessary.
Group Balance is an activity that requires group formation and recognition of personal roles. The game is played using cooperation to reach the common goal. In the process, group members create interpersonal connections and learn from their mistakes to complete the tasks. In the reflection section, participants first report on their concrete experiences, such as feelings about their personal actions that influenced the group goals, pressure, and support from group members, and identify the person who acted as the leader. Regarding the practice, the participants are asked to describe the attitude and method they would use to coordinate a group and communicate with group members in similar experiences in the future, and to describe their thoughts regarding the interrelationship between partners and the actions they would take to establish a balance. At the end of the game, the following feedback is provided.

*The board in the game represents trust in human interaction. Trust and support are complementary. Clear goals and member complements hold a team together. Long-term trust can suddenly be lost because of discreet insensitivity. Arguments occur, and are a group dynamic. How arguments are dealt with influences group creativity. A positive cycle of strategies, thoughts, communication, and attitude enables the group to succeed.*

**Method**

Before the research commenced, Delphi method was adopted to generate adventure education themes. Eight experts who had academic or industrial backgrounds and abundant experience in teaching or conducting professional practice in the field of adventure education were invited to participate in the exploration process by completing questionnaires. Twelve adventure education activities that required participants to avoid using one of their senses, such as speech or sight, were analyzed. The basic qualities of such activities are appropriate for digitization. After three-round discussions, 13 common themes were generated and classified into two primary sections: extrinsic behaviors and intrinsic motivations. The extrinsic behavior domain comprised leadership, communication, collaboration, support, responsibility, and activeness, and the intrinsic motivation domain comprised reflection, trust, creative thinking, empathy, willingness to change, acceptance of frustration and failure, and confidence. The research procedure used in this study is shown in Figure 7.

Pretest and posttest questionnaires were designed to understand the opinions of students regarding the game system and the effectiveness of adventure education. The questionnaire covered the 13 adventure education themes, each of which was assessed using two items, ranked on a 5-point Likert scale, yielding a total of 26 questions. Questions, for example, for communication aspect was “I thought good communication between group members can help members to understand each other and solve problems together” for pretest, and “I thought good communications between members can help them solve problems” for the posttest. The reliability of the
questionnaire was high, exhibiting a Cronbach’s α of .70 and a coefficient of internal consistency of .90. Given that there were 13 themes with 2 questions per theme, the Likert scores on each theme were averaged and 13 pre-post t-tests run statistically. Then original scores were compared to view minor differences between the pretest and posttest of each theme with qualitative descriptions.

A total of 40 college students, aged between 18 and 22, were invited to participate in the activities. They were equally distributed into two groups, a physical group and a virtual group. In the physical group, the adventure education games were conducted in person and guided by a facilitator. In the virtual group, the adventure education games were conducted physically with digital games and guided by a virtual facilitator predesigned in the system. In the process, students played the game with group members, solved tasks, and reflected on the questions generated by the system concerning the adventure education themes.

The activity processes were recorded in the system. Six researchers conducted observations during the process to investigate the engagement of the participants and their achievements in adventure education. Reflection sections were conducted after each game. After the students completed both games, researchers interviewed them and analyzed the dialogue to determine the effectiveness of the games. Finally, game satisfaction was evaluated to determine participants’ regarding about the digital games.

Results

Players’ achievements in learning the adventure education themes

The pretest and posttest questionnaire results regarding the players’ achievements in learning the adventure education themes (Table 1) were summed up to test the differences. It is found that the t-test score of the physical group ($t = 5.16, p = .000$) and the virtual group ($t = 3.61, p = .000$) both reached significant differences. Both the digital adventure education games and the physical adventure education activities effectively enhanced the participants’ conceptual understanding of the adventure education themes.

![Figure 7. Research procedure](image_url)

### Table 1. Pretests and posttests of players’ achievements to the adventure education themes ($n = 20$)

<table>
<thead>
<tr>
<th>Group</th>
<th>Pretest Mean</th>
<th>Pretest SD</th>
<th>Posttest Mean</th>
<th>Posttest SD</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical</td>
<td>4.28</td>
<td>.333</td>
<td>4.72</td>
<td>.514</td>
<td>5.16***</td>
</tr>
<tr>
<td>Virtual</td>
<td>4.38</td>
<td>.320</td>
<td>4.83</td>
<td>.439</td>
<td>3.61**</td>
</tr>
</tbody>
</table>

Note. *$p < .05$, **$p < .01$, ***$p < .001$.

Analysis of both themes indicated a slight difference between the pretest and posttest scores of the two groups.

Regarding the extrinsic behavior domain, three themes, namely leadership, cooperation, and activeness, exhibited a positive differences (posttest scores higher than pretest scores) in both groups, signifying that both digital games and physical activities effectively enable participants to achieve the abilities addressed in these themes.
Conversely, three themes, namely communication, support, and responsibility, exhibited lower scores of posttest scores to pretest scores in the virtual group, signifying those participants in the virtual games placed more attention on the completion of tasks and challenges and were less aware of the formation of social groups. Therefore, they sensed less personal responsibility in the groups, less communication occurred between group members, and the members showed less support to each other. By contrast, the physical group exhibited higher posttest scores to pretest scores in these themes because the eye contact and sense of physical existence involved in face-to-face human interaction can improve group support and communication. Participants were more likely to identify with individual or group responsibilities during the activities in situations of direct human contact than in situations of indirect contact.

| Table 2. Players’ achievements to the adventure education themes: Extrinsic behaviors |
|---------------------------------------------|---------------------|---------------------|
| Themes                          | Physical group | Virtual group |
|                               | Pretest | Posttest | Difference between means | Pretest | Posttest | Difference between means |
| Leadership                     | 3.78   | 3.9     | 0.12                  | 3.71   | 4.27     | 0.56                  |
| Communications                 | 5.1    | 5.25    | 0.15                  | 5.19   | 5.08     | -0.11                 |
| Cooperation                    | 4.98   | 5.03    | 0.05                  | 4.73   | 4.97     | 0.24                  |
| Support                        | 4.85   | 4.98    | 0.13                  | 5.07   | 4.99     | -0.08                 |
| Responsibility                 | 4.98   | 4.50    | 0.20                  | 5.04   | 4.09     | -0.14                 |
| Active                         | 4.7    | 4.83    | 0.13                  | 4.62   | 4.72     | 0.10                  |

In the extrinsic behavior domain, the average of the posttest scores was higher than the average of the pretest scores for both groups in the leadership, cooperation, and activeness themes, indicating that the digital adventure education game positively affected these themes.

Leadership refers to how an individual leads a group toward completing a task when the group encounters difficulties. The results indicate that the participants actively participated in the game and exhibited clear behavioral differences regarding showing leadership. The virtual group exhibited a greater difference than the physical group did, indicating that the participants were likely to hide behind technology and reveal their true selves and behaviors when interacting through the digital game. However, in the physical group, the participants actively tried to improve their self-images, causing them to hide parts of their inner selves from the others and change their responses according to the group atmosphere.

Cooperation refers to how the participants cooperate with other members to achieve the group goal. The results indicate that the groups achieved goals more effectively when the participants cooperated instead of worked individually. The virtual group exhibited a greater learning effect, possibly because the digital games had higher failure rates than the physical activities, causing members to work closely with each other to avoid restarting the game. Therefore, the participants in the virtual group were more aware of the need to cooperate than were those in the physical group.

Activeness refers to how the participants actively confronted the problems in the game and solved them to achieve the group goal. The results indicate that the digital games exhibited similar effectiveness in motivating participants to actively participate the activities. The physical group exhibited more favorable performance in this theme because these participants experienced greater pressure to construct personal images and improve interpersonal relationships.

Regarding the communication, support, and responsibility themes in the extrinsic behavior domain, the significant differences observed in the physical group were greater than those observed in the virtual group. Communication refers to how the participants communicated with each other when encountering difficulties. Communication can enhance interpersonal relationships and establish favorable group cooperation. The results indicate that in the virtual group, the posttest score ($m = 5.08$) was lower than the pretest score ($m = 5.19$), possibly because the participants in the physical group tended to communicate with each other to generate solutions. They approached the games slowly and listened to each other before taking action. However, to complete the tasks in the digital games, the participants felt little need to communicate. In addition, the participants seemed eager to finish the task so that communication did not occur often; this was ineffective. Furthermore, the participants in the virtual group tended to work alone.

Support refers to how the participants provided fellow group members with support and worked toward the goal collectively. The results also indicate that in the virtual group, the posttest score ($m = 4.99$) was lower than the pretest score ($m = 5.07$).
Responsibility refers to how the participants helped fellow group members grow together and regarded themselves as crucial members in the group. The results indicate that in the virtual group, the posttest score ($m = 4.09$) was lower than the pretest score ($m = 5.04$). This is because the digital game participants did not see the relationship between themselves and the group, and regarded the game tasks as their personal responsibilities. They hesitated to provide others with support. However, in the physical group, the participants interacted directly, stimulating interpersonal interaction.

Overall, the study revealed that the physical group exhibited significant differences between the pretest and posttest scores in every theme of the intrinsic behavior domain. However, in the virtual group, communication, support, and responsibility were less observable than they were in the physical group. The posttest scores were lower than pretest scores, indicating that digital games cannot effectively increase participants’ communication ability and reduce their ability to show support to others.

Regarding intrinsic motivation, four themes, namely reflection, creative thinking, accepting frustrations and failures, and confidence, the posttest scores were higher than pretest scores in both groups, signifying that both digital games and physical activities effectively enable participants to achieve these abilities.

Three themes, namely trust, empathy, and willingness to change, the posttest scores were lower than pretest scores in the physical group, indicating that individuals’ internal feelings and thoughts were less evident in situations of direct human contact than in situations of indirect contact. By contrast, because the digital environment creates a sense of private space, people can easily exhibit their care, concern, empathy, and willingness to change without worrying about face.

In the intrinsic motivation domain, the average posttest score was higher than the average pretest score of both groups in the reflection, creative thinking, acceptance of frustration and failure, and confidence themes, indicating that digital adventure education games are similarly effective or more effective compared with physical activities.

<table>
<thead>
<tr>
<th>Project</th>
<th>Physical group</th>
<th>Virtual group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reflection</td>
<td>4.45</td>
<td>4.54</td>
</tr>
<tr>
<td>Trust</td>
<td>4.75</td>
<td>4.42</td>
</tr>
<tr>
<td>Creative Thinking</td>
<td>4.45</td>
<td>4.52</td>
</tr>
<tr>
<td>Empathy</td>
<td>5.2</td>
<td>4.83</td>
</tr>
<tr>
<td>Willing To Change</td>
<td>5.08</td>
<td>4.54</td>
</tr>
<tr>
<td>Accept Frustrations and Failures</td>
<td>3.63</td>
<td>3.68</td>
</tr>
<tr>
<td>Confidence</td>
<td>4.28</td>
<td>4.43</td>
</tr>
</tbody>
</table>

In the intrinsic motivation domain, the average posttest score was higher than the average pretest score of both groups in the reflection, creative thinking, acceptance of frustration and failure, and confidence themes, indicating that digital adventure education games are similarly effective or more effective compared with physical activities.

Reflection refers to how the participants answered their questions through reflection, personal learning, and growth. The results indicate that both groups exhibited a similar understanding of this theme. The physical group exhibited more favorable performance in this theme than the digital group did because the facilitator led the discussion by using in-depth observations and established techniques. Although the virtual group received the same questions as the physical group did, the participants directly responded to the questions without being instructed to elaborate or clarify.

Trust refers to how participants believed in other group members to offer support and assistance. The results indicate that the virtual group exhibited clearer differences than the physical group did; this may have been because the participants tended to hide behind technology rather than provide assistance to others. They felt safe and relaxed, and were not concerned with embarrassment, which frequently occurs in activities involving direct human contact.

Creative thinking refers to how participants determine solutions by using unconventional thinking models and frameworks. The results indicate that the physical group exhibited more favorable performance in this theme than the virtual group did because face-to-face situations stimulate communication and creative thoughts more than the lagged discussion in digital games does.

Acceptance of frustrations and failure refers to how the participants remained willing to face challenges when failure occurred. The results indicate that the participants in the virtual group exhibited a higher acceptance level
than did those in the physical group, possibly because they generally play the games from the third-person perspective and feel more distant from the direct consequences of failure.

Confidence refers to how the participants collaborated with group members to solve problems and increase their self-confidence. The results indicate that the virtual and physical groups exhibited a similar group atmosphere; specifically, they both completed tasks collectively and gained confidence after completing the tasks.

Regarding intrinsic motivation, the virtual group exhibited more favorable performance than the physical group did in several themes. Empathy refers to how the participants considered other members’ needs. In the physical group, the pretest score \( (m = 5.20) \) was higher than the posttest score \( (m = 4.98) \). In the virtual group, the posttest score \( (m = 5.00) \) was higher than the pretest score \( (m = 4.83) \). These results indicated that empathy is an inner thought that can be more easily revealed in digital space than in the face-to-face situations. Because of shyness and embarrassment, the digital space provides participants with a safe place in which inner thoughts can be revealed easily.

Willingness to change refers to how the participants regarded making changes and generating new ideas as a vital factor to self-development. In the physical group, the pretest score \( (m = 5.08) \) was higher than the posttest score \( (m = 4.73) \). In the virtual group, the posttest score \( (m = 4.79) \) was higher than the pretest score \( (m = 4.54) \). In digital games, innovation seems to be an essential characteristic.

**Satisfaction with the digital adventure education game**

Digital tools were used to develop an adventure education game to create a learning experience similar to traditional face-to-face activities. The design of the game system and content both influenced the participants’ gaming experience and affected their learning effectiveness. Therefore, the satisfaction questionnaire investigated whether the game successfully conveyed the adventure education characteristics and goals by recording the participants’ views and game satisfaction.

The questionnaire consisted of eight questions. The overall system evaluation average score was high \( (m = 4.59) \). Challenge exhibited the highest score \( (m = 4.85) \), indicating that the game contained appropriate difficulty levels and challenges. The participants solved problems, thereby developing their problem-solving abilities. In addition, their reflection ability was enhanced after playing the digital adventure education game \( (m = 4.73) \). All participants could easily manipulate the game according to their personal needs \( (m = 4.60) \). The lowest score was the interaction score \( (m = 4.40) \), indicating that participants mostly undertook personal challenges while playing the game. The immersion score \( (m = 4.50) \) indicated that the scenarios of the game can be improved.

**Discussion and conclusion**

Meta-learning is a major focus of this study. In the game, the reflection section was designed to generate synthesis feedback after the game.

In the Corporate Maze reflections, participants felt that the most difficult aspects of the game were determining the next step and experiencing doubt regarding the choice. Exploring the unknown made them nervous, but the process encouraged them to make brave attempts. Group members needed to remember the correct route even though they were not the player. In addition, the participants learned that “breaking the frame” is crucial. Creativity and innovation were necessary to take steps toward the goal. Such lessons are similar to making life choices while facing the unknown.

In the Group Balance game reflections, the participants reported that the most essential aspects of cooperation are personal roles, nurturing a tacit sense of connection, and observing others’ needs. They needed to encourage instead of blame each other when failing the task and stay calm during the challenges. Only communication could enable them to successfully balance the board. While they played the game, they learned the importance of cooperation and communication, and that a leader is necessary to mediate the communication. During the process, most leaders were those standing in the center and controlling the board. Most recorded sounds were the screams of group members and directions from the leaders.

The goal of this research was to encourage the participants to explore themselves through the adventure education activities and to enhance their interactions with other people. In the digital game, a community was developed in
which the participants could communicate, cooperate, achieve group goals, and establish self-confidence. Through digital tools, the learning space was extended to the virtual space which improves freedom and comfort. Using the digital games, facilitators and researchers can record the gaming process, provide information, and assist the participants in conducting cognitive reconstruction. They can solve problems and modify their behaviors. The participants learned strategies that can be used in the future, and their self-efficacy improved.

Research results have indicated that teenagers experienced a greater sense of belonging to the groups and identification with digital games. Adventure education is a series of educational activities that use the physical senses. Reflection sessions are the core of adventure education, and are used to form group identities. Every group task requires problem-solving skills, the use of which causes individuals to appreciate the importance of personal responsibility to the group. This is the practice of Kolb’s experiential learning.

Digital adventure education games focus on group cooperation and interpersonal interaction; therefore, the design should begin with educational applications and can be developed further to become an independent teaching module used by school teachers as classroom activity tools. Teachers can thus enrich the teaching content and enable students to develop positive thinking and strengthen their skills.

From the observation and interview results, the digital game provides a safe distance for the students between their self and the sub-consciousness, and functions as a communication means for social interaction. It enables learners to experience and feel themselves in the gaming process, reflect on their inner characteristics, fully express and reveal their emotions, ideas, experiences, and behaviors, which foster the strength to change and mature.

Through multiplayer digital games, players can reflect on key elements such as teamwork, trust, responsibility, problem solving, decision making, leadership, goal setting, and communication, because the games are embedded with metaphorical teachings, interpersonal communication, self-involvement, game immersion, and emotional liberation functions.

Learning using digital games can motivate learners more than traditional learning can. In this study, compared with traditional learning, more knowledge can be gained and learning transfer was stronger and more predictable in learning using digital games.

In traditional adventure education activities, facilitators cannot observe each participant and record the complete process of every game. Facilitators can only reflect on specific interaction events, reply to group interviews, and provide feedback and guidance based on their feelings at the time. Precision in individualistic feedback is therefore difficult. However, digital game system can represent the process of physical adventure education activities and involves providing personal feedback. The gaming experience can be analyzed using learning assessments and problem identification methods. In addition, motion-sensing games emphasize metacognitive development through the discussion of metaphors that are helpful and essential for learning achievement.

Adventure education activities have diversified characteristics and values; they are widely used in various curricula. However, digital technology has not been integrated because personal experience is its most crucial element. This study exploited the simplicity and diversity of digital technology to complement the limitations of physical activities. In the transformation from a physical environment to a virtual environment, personal experience, fun, and reflections are retained in the activities when users are immersed in a game. Through multiplayer digital games, players can reflect on key elements such as teamwork, trust, responsibility, problem solving, decision making, leadership, goal setting, and communication, because digital games are embedded with metaphorical teachings, interpersonal communication, self-involvement, game immersion, and emotion liberation functions. It is evident in this study that it is effective to advance adventure education activities with digital game, especially motion-sensing games.

Acknowledgments

This study is supported in part by the Ministry of Science and Technology (previously known as National Science Council) of the Republic of China, under contract numbers NSC 100-2628-S-024-002-MY3 and NSC 101-2511-S-024-009-MY3.
References


Evaluation of Intercultural Instructional Multimedia Material on Implicit Xenophobic Cognition: Short Time Effects on Implicit Information Processing

Joerg Zumbach1*, Gerhard Schrangl1, Chad Mortensen2, and Stephanie Moser1

1Department of E-Learning in Science Education, University of Salzburg, Salzburg, Austria // 2Department of Social and Behavioral Sciences, Midlands Technical College, Columbia, USA // joerg.zumbach@sbg.ac.at // gerhard.schrangl@gmail.com // mortensenc@midlandstech.edu // stephanie.moser@sbg.ac.at

*Corresponding author

(Submitted July 8, 2015; Revised November 17, 2015; Accepted January 15, 2016)

ABSTRACT
Considering xenophobic attacks against foreigners and ethnic or religious motivated wars, there is a need for educational concepts to extinguish xenophobia. A model describing the cognitive processes involved in Xenophobic cognition was developed. Instructional multimedia material that discussed various forms of alienation was developed and evaluated. The computer program was based on research findings and proven intercultural teaching strategies. To evaluate the training sessions, methods were developed to measure attitudes toward foreigners. Based on a multi-method concept an evaluation-tool was created that validated the effectiveness of long-term intervention programs and short-term effects on an implicit level of awareness. The procedure used the implicit association test. The applied methods were evaluated in a study with 79 High School students (46.8% female; mean age = 14.13, SD = 1.16) within a one-factorial control group design. Although there were no significant results in explicit measures of attitude, there was a significant change in implicit measures showing a reduction of implicit xenophobic cognitive processes in the experimental group and an increase in the control group. Results imply that the chosen tutorial methods lead to a change in attitudes on trait and state-level regarding xenophobic self-concept.

Keywords
Social cognition, Implicit association test, Intercultural education, Prejudice

Introduction

Understanding another’s cultural backgrounds as well as our own is one of the most important skills in today’s society. Every day, we face situations and social interactions that might be perceived as strange because they are different from our own culture. However, intercultural education could support the process of establishing a peaceful coexistence between cultures when one is confronted with new and unfamiliar intercultural exchanges. That is, intercultural education might contribute to support a change of perspective that can reduce prejudices, resentments, and unfounded fears of any type of foreignness one may encounter.

Most of our beliefs and ideas of other countries and cultures are not based on personal contacts with foreign people but rather are outcomes of traditional lore (cf. Budke, 2008). Such tradition might manifest itself in xenophobic jokes, expressions, cartoons, and propaganda of ultraconservative and right-wing parties. Stereotypes might lead to dubious conclusions and predictions about people from foreign countries. Sometimes, even teaching (e.g., geography courses), might contribute to the generation of stereotypes (cf. Markom & Weinhäupl, 2007).

However, demographic change makes immigration necessary, for example, to sustain the Western socio-economic system. As such, intercultural competencies, and thus, intercultural education is necessary in a multicultural society. Globalization leads to the fact that people are confronted with foreignness within their traditional settings. To avoid a “Clash of Civilizations” as predicted by Huntington (1993), and to assure that ethnic diversity will be recognized as an economic and social viability, a specialized education seems to be necessary. This education should serve to mediate cultural differences among people of different origins.

The current investigation was designed to develop and examine a specialized instructional multimedia material that was based on recent psychological and pedagogical research findings. The main objective of the learning program was to foster attitudinal change in students regarding prejudices, stereotypes, and perceptions of foreignness against people from other cultures. A higher level of education is often seen as a predictor of tolerance and positive attitudes towards a multicultural society. Education also enhances tolerance for others (cf. Noack, 2001; Rippl & Seipel, 2002). Pettigrew and Martin (1987) suggest that education might contribute to a change of attitudes and, in turn, help reduce xenophobic attitudes. Therefore, a learning environment was
developed that could help learners’ reflect upon their behaviors and attitudes towards foreigners and consequently, reduce the xenophobic attitudes they may have held. Nevertheless, conducting such training does not automatically guarantee its success. It still remains unclear if education can lead to a change in xenophobic attitudes. Hence it is necessary to evaluate the effectiveness of the training. In the present study, the success of the learning program depended on the strength of the negative association to foreignness and the ability of the intervention program to produce a change. Negative associations are sometimes difficult to change because they are products of socialization and educational processes that have developed over many years.

Intertwined with this problem is the finding that assessing prejudices, stereotypes or other dimensions of xenophobia implies the use of a multi-methodological approach, to obtain not only explicit, but also covert, implicit values. Thus, as part of the current empirical investigation, several methods were developed to measure xenophobic attitudes and stereotypes.

Taken together, the current study was designed to answer the following research question: can an instructional multimedia material lead to a significant change in attitudes as well as covert and implicit xenophobic self-concepts?

The following behavioral learning approach utilized in the study was based on a general model to explain xenophobic behavior and attitudinal change.

A General Xenophbic Model

Implicit prejudiced attitudes are introspectively unidentified traces of past experiences (e.g., Greenwald & Banaji, 1995). The General Xenophobic Model (GXM) describes the concrete development of xenophobic cognitions that are based on general explicit and implicit learning models (e.g., the General Aggressive Model; cf. Anderson & Bushman, 2001). The basic assumption of the GXM (see Figure 1) is that there is a connection between observation, socialization, and concrete xenophobic behavior (cf. Schragl & Zumbach, 2012).

The GXM has been designed to form a framework for understanding the cognitive processes that are based on repeated observations that lead to xenophobic attitudes. The model also considers implicit cognitive processes and integrates the dual process model by Devine (1989). It assumes that there are automatic and controlled components involved in the development of stereotypes.

Figure 1. Short-term processes of the General Xenophobic Model
The model shows the importance of observation regarding the development of xenophobic behavior. The observation can be realized through direct contact (primary experience) or communication (secondary contact; Müller, 2004).

The model describes the processes of how people acquire xenophobic memory representations. Within a specific situation, a person is confronted with one or more persons from a different culture. During the situation, a person experiences a specific state (e.g., frustration) that acts as an input in the subsequent information processing. The processing of the input variables can be done on an affective, cognitive, or arousal level. Let us assume that a person is frustrated because he or she has just lost his or her job. He or she watches a TV advertisement of a right-wing party where unemployment is attributed to the high number of immigrants that take jobs from nationals. The person may process the input cognitively and, thus, might internalize this causal relation. Additionally, the person might also react affectively, for example, becoming increasingly angry towards immigrants. Furthermore, the affective reaction might also influence other accompanying processes that trigger physiological arousal, such as an increased heart rate or blood pressure. The information processing results in an evaluation of the situation that might lead to an explicit and thoughtful reaction (e.g., actively voting for the party) and/or an implicit and rather impulsive behavior (e.g., fostering stereotypic thinking against immigrants) that might impact subsequent social interactions. If the xenophobic behavior does not lead to negative, but rather neutral or positive consequences for the person, the probability of acquiring xenophobic attitudes, behaviors, and knowledge structures increases, and might lead to new, similar information processing and evaluation processes. Thus, starting a new cycle that supports xenophobic attitudes. It is also important to note that the situation does not have to necessarily be experienced in the real world, but can also be mediated by media (cf. Esser, Scheufele & Brosius, 2002).

Repeated consumption of prejudical media and observation of racist behavior might contribute to the development of xenophobic knowledge structures (see Figure 2). Such structures are usually complex and very difficult to change. The xenophobic knowledge structures also lead to consious (explicit) and impulsive (implicit) decision making (cf. Devine, 1989) in corresponding social interactions (e.g., Berkowitz, 1993). The first automatic reaction and the following controlled judgement of the person lead to the production of specific behaviors. People with a strong tendency toward xenophobic behavior can easily access xenophobic behavioral scripts due to social priming situations. Support is given here by Greenberg and Pyszczynski (1985) who were able to show that racist statements easily activated stereotypes and, thus, lead to discriminatory acts.

![Figure 2. Long-term processes of the General Xenophobic Model](image-url)

Repeated experience of the short-time model cycle might contribute to permanent long-term effects. These consequences are illustrated in Figure 2. Long-term learning processes are triggered by repeated observations of xenophobic behaviors. Possible consequences include xenophobic beliefs and attitudes (cf. Bacher, 2001), xenophobic patterns of perception (Wahl, 2003), prejudicial behavioral scripts (Greenberg & Pyszczynski, 1985), desensitization towards xenophobic statements, and a growing detachment towards foreigners (Wahl, 1999). All these effects are caused through exposition, socialization, and observational learning.
Xenophobic attitudes can be a result of socialization (Bacher, 2001). While Wahl (2003) assumed that xenophobic attitudes were the product of evolutionary processes, the current investigation suggests that it is a consequence of observational learning experiences that lead to negatively associated stereotypes against foreigners from different cultures. Additionally, people that have already developed xenophobic tendencies often avoid contact with foreigners. However, contact with foreigners might actually reduce xenophobic attitudes. According to the contact-hypothesis theory, intercultural contact has been shown to have a positive impact on attitudes towards foreigners (cf. Allport, 1954). Furthermore, the process of self-reflection provides one a means of changing one’s own attitudes (Gudjons, Pieper & Wagener; 1986).

Taken together, the suggested model assumes that each single xenophobic episode can contribute to the construction and activation of xenophobic knowledge structures. Such structures might contain perception failures such as illusory correlations (cf. Shavitt, Sanbonmatsu, Smittipatana & Posavac, 1999) or hostile expectations (cf. Anderson & Bushman, 2001). Repeated experience of the cycle might also lead to desensitization towards hostile behavior against people from other cultures. In short, the above factors are likely to contribute to the adoption of a xenophobic personality—in a similar to fashion to how aggressive behaviors are learned (cf. Anderson & Bushman, 2002).

For example, Huesmann (1994) found a strong connection between socialization and personality development. While socialization often takes place within families, the contact that people have outside of their own family can also have an impact on personality development. A person’s peergroup and other adults (e.g., teachers or trainers) also play an important role (cf. Huesmann, 1994). If the peergroup has a high acceptance of xenophobia, then one may tend to find the behavior more acceptable. Furthermore, non-xenophobic peers tend to be rejected by members of xenophobic peers, whereas, like-minded people tend to attract each other.

It is therefore assumed that the kind of learning described in the model above (e.g., observational learning) can significantly influence the perception of strangers and foreignness.

**Explicit and implicit attitudes**

As described above, the model considers implicit cognitive processes as well as automatic and controlled components involved in the development of stereotypes. Explicit cognitions are introspectively accessible, while implicit processes operate outside of awareness. Contrary to explicit beliefs, implicit attitudes are introspectively unidentified traces of past experiences that mediate favorable or unfavorable feelings, thoughts, or actions toward social objects (Greenwald & Banaji, 1995), as well as influence automatic and spontaneous behavior (Gawronski, 2006; Gschwendner, Hofmann & Schmitt, 2006). In other words, it is possible to be prejudiced without making a conscious decision.

While explicit measures mostly rely on self-reports, implicit measures are frequently based on reaction time tasks, such as the Implicit Association Test (IAT, Greenwald, McGhee & Schwartz, 1998). Generally it is possible that there is a divergence between explicit and implicit measurement results, particularly regarding socially sensitive topics like prejudice against minorities (Hofmann, Gawronski, Schwendner, Le & Schmitt, 2005). Implicit attitudes can be seen as better predictors of behavior than self-reported attitudes because a participant might lack the ability to introspect correctly (Brunel, Tietje & Greenwald, 2004). Similarly, another problem with explicit measures is that participants might be motivated to avoid external censure by answering questions in a manner that will be viewed as “good behavior” by others. Thus, socially desired answering might be a consequence. By using reaction time measures and excluding responses that take too long, socially desired answering is minimized (Gschwendner et al., 2006). The main difference between explicit and implicit measurement methods is that explicit methods allow an introspective approach of the measured constructs, whereas, implicit methods do not. Thus, by detecting spontaneous processes, the implicit attitude can provide a more comprehensive picture of a person’s attitudes.

The first question is whether or not implicit and explicit decisions can be influenced by a short-term instructional intervention that helps one reflect about situational or personal input variables? The second question is how to measure the influence of instructional approaches on implicit knowledge structures? We focussed on an implicit, cognitive level rather than on a genuine affective level because it is assumed that implicit cognitive structures would subsequently have impact on affect. The present study will address both of these questions by evaluating an intercultural training program with established explicit measures (questionnaires) as well as a proprietary Implicit Association Test (IAT) to assess changes in implicit knowledge structures related to the perception foreignness. Since short time interventions rarely lead to significant changes of explicit attitudes, a
combination of explicit and implicit measures was used to better detect the initial implicit changes that subsequently affect explicit changes (e.g., Cochran-Smith, 2003; McDiarmid & Price, 1993; Wasonga, 2005).

Method

The present study describes the development, application, and assessment of an intercultural training software program as well as a methodological toolkit to assess short- and long-term consequences of instructional interventions for reducing xenophobic attitudes. The study employed an experimental-control group, one-factorial design with repeated measurement. The IAT was administered before and after the intercultural teaching sequences. Additionally, explicit measurements for operationalization of emotional and cognitive effects were also used.

Participants

The participants were 79 Austrian high school students (46.8% female) between the ages of 12 and 16 ($M = 14.13$, $SD = 1.16$). Sixty-one participants were assigned to the experimental condition, 18 participants were assigned to the control group. Due to administrative processes the control group was assessed half a year later in a comparable school within the same class levels. Thus, the original planned experimental design had to be modified to a quasi-experimental design. Participation was voluntary with no incentives and was completed during regular school hours. High School students were selected based on the rationale that this group is highly vulnerable for stereotypes and prejudices, but can highly benefit from programs to overcome them. Participants were grouped in a manner consistent with other programs such as the group jigsaw (Aronson & Patnoe, 2011).

Material

The Implicit Association Test

To detect participants’ implicit attitudes we used the Implicit Association Test ($\alpha = .79$; Greenwald, McGhee & Schwartz, 1998). The IAT detects the strength of association between a person’s different mental concepts. The test is based on scientific research concerning implicit social cognitions (Greenwald & Banaji, 1995).

The IAT measures individual differences in automatic activation of semantic and evaluative associations (Gawronski & Conrey, 2004). It is a reaction-time-based measurement method of automatic association strength between two concepts. In short, the IAT makes it possible to measure implicit cognitions.

Theoretical background of the IAT

The IAT-effect results from different reaction times while assigning stimuli to dichotomous categories. According to the Activation Theory of Semantic Processing, the difference in reaction times represents the concept of spreading activation. (Collins & Loftus, 1975). It assumes that when concepts in memory are activated, those concepts that have been learned or experienced in conjunction with the activated concept will also be activated. Thus, concepts become more or less connected to each other.

In the IAT-task, the subject has to categorize various stimuli into dichotomous concepts. Faster reaction times indicate stronger associations. Congruent tasks can be solved faster than association- incongruent tasks (Greenwald et al., 1998). Since the IAT uses complementary pairs of concepts, it only measures the relative strength of pairs of associations (Greenwald & Farnham, 2000). Thus, this method is not measuring the strength of associative links of memories, but the difference between various associative links (Greenwald & Nosek, 2001).

Evaluation of the visual material of the IAT

Appropriate visual stimuli play a major role during the IAT process (Nosek, Greenwald & Banaji, 2007). Therefore, the visual materials of the IAT had to be created and evaluated before its application. The visual materials used in the study were taken from the International Affective Picture System – IAPS (Lang, Ohman &
Vaitl, 1988) and from daily life scenarios in Austria. To get dichotomous categories for the IAT, the visual materials were evaluated according to the degree of familiarity. Eighty participants (47.2% female) between the ages of 10 and 76 (M = 19.95; SD = 9.75) evaluated a set of pictures using a semantic differential scale ranging from culturally distant to culturally proximal. The sample consisted of Austrian high school students (61.2%), university students (31.2%), and other adults (7.6%).

The data was subjected to a principal component factor analysis with varimax rotation (KMO = .500; Bartlett’s = .005). The pictures were divided into familiar and non-familiar categories. The most common attribute of the respective groups was skin color.

All pictures showing dark-skinned people were assigned to the non-familiar group (see Figure 3) and all pictures showing light-skinned people were assigned to the familiar group (see Figure 4). The pictures were integrated into the IAT as dichotomous categories—familiar and unfamiliar.

**Figure 3. Cultural distant pictures**

**Figure 4. Cultural proximal pictures**

### The IAT procedure

The IAT consists of five discrimination tasks where participants are required to respond as quickly as possible. Participants complete an associated attribute discrimination task where the participants are asked to differentiate between pleasant and unpleasant words and pictures (e.g., health, heaven, poison, murder; see also Nosek et al., 2007). In the third task, the two prior tasks are superimposed or mapped onto one another. For example, one response key is used for individual stimuli that are either familiar or pleasant. Another response key is used for stimuli that are unfamiliar or unpleasant. In the fourth step, the response keys are reversed. Finally, in the fifth step, the mapping from the stimuli is reversed from the third step. Individual IAT effects were computed as D-scores (Greenwald, Nosek, & Banaji, 2003). That is, differences between the mean latencies in the critical blocks 3 and 5 were calculated by an individual’s pooled standard deviation score. Higher scores indicate higher xenophobic attitudes. Most participants had negative IAT scores. However, given a fixed block order and the absence of an IAT calibration procedure of zero scores, it was important that the absolute magnitude of the IAT effects were not overinterpreted (see also Zumbach, Seitz & Bluemke, 2015).

### Explicit measurements

We used several self-report measures to determine participants’ attitudes. The first scale was the “Motivation to Unbiased Behavior” questionnaire by Banse and Gawronski (2003). It is divided into three subscales “Behavioral Control” (8 items; e.g., “One should not laugh at jokes about foreigners”; α = .80), “Admission of Prejudice” (4 items, e.g., “I pay attention to not let prejudices influence my behavior”; α = .66) and
“Unprejudiced Self-perception (4 items, e.g., “I’d feel uncomfortable if someone believes that I have prejudices toward minorities”; \( \alpha = .51 \)).

Second, we used the “Xenophobia” scale (Bucher, Göllner & Auer, 2001; \( \alpha = .73 \)) with seven items, e.g., “There would be much less problems in our country if less foreigners lived here.”

Finally, we used the scale “Attitude Towards Foreigners” by Frindte, Funke und Jacob (1999). It detects participants’ position regarding foreigners by means of nine items (e.g., “Foreigners provoke xenophobia by their own behavior”; \( \alpha = .85 \)).

Participants indicated their answers to all items using a 5-point Likert scale (1 = “I do not agree at all”; 5 = “I do totally agree”).

**Educational software “Intercultural Games”**

The main goal of the study was to have students in the experimental group engage with instructional multimedia material, designed to reduce xenophobic attitudes and schemata. The “Intercultural Games” consisted of several modules that helped learners become familiar with elements of foreignness, to accept culturally distant stimuli, and to reframe them as a part of everyday life. The program is based on a biographical approach to the topic of foreignness. It uses “biographical self-reflection” as well as an analytical and self-reflective process.

Gudjons, Pieper and Wagener (1986) define a biographical self-reflection as a reappraisal of one’s own biography. Past experiences shape identity and behavior. Through the process of reappraisal we can explain and change our own attitudes. A dialectical entanglement of the reflection and self-reflection must always be part of intercultural education (Zielke, Meier & Bollacher, 2005). Self-reflection is an important aspect of multicultural education whereby discriminations can be made visible (Hejazi, 2009). For this reason, the tasks of the tool addressed participants’ biography and the life of strangers by focusing on the encounter between people of foreign origins.

Analyzing the images and films were an essential part of this approach. Furthermore, the intervention tool was based on different social and educational theories including symbolic interactionism (Mead, 1978) and symbol-didactics (Biehl, 1991). In everyday life, symbols play an important role in communication, identity formation, and also localization in social space (cf. Holzwarth 2001). In short, the training program was designed to not only allow students to come into contact with foreigners, but to also allow them to discuss and become familiar with various types of alienation.

**Music Example I**

Please start the music by clicking on the symbol below the map. After listening please click on the cd-symbol button and pull it to the eventual music producing country in the map. Release the mouse button! If you can’t find the country, you can click on the „I don’t know“-button.

![Figure 5. Screenshot from module “music from around the world”](image-url)
The instructional multimedia material includes seven different modules. The first module, “music from around the world,” requested participants to listen to several pieces of music first and relate it to the corresponding parts of the world afterwards (see Figure 5). The diverse styles of music varied from vocals of several ethnos to traditional folk music. During the module “viewing pictures,” pictures showing different cultural habits were presented along with additional information, e.g., women wearing scarfs in religious tradition (see Figure 6). Afterwards, learners had to answer topic-related questions. In the “symbols of foreignness” module, participants had to arrange cultural and religious symbols according to familiarity, e.g., the Star of David or the Yin and Yang sign. Additionally, learners could retrieve audio-visual information related to the symbols.
The “create a crest” module, was designed to help learners through a process of self-discovery (see Figure 7). In this module, learners had to create a crest that is related to their own identity. Several cultural and religious symbols were presented as well as shapes of different countries and continents. Learners chose symbols and arranged them within a pattern of their choosing. The individual crest was printed and discussed in class afterwards. During “foreign pictures” module, participants had to arrange photos. The pictures were taken of persons and things in the immediate vicinity of the learners. Photos could be classified either as foreign or familiar. Subsequently, an explanation was presented to that dealt with the term of “foreignness.” It allowed a better understanding of the feeling that it can be hard to decide whether a picture is familiar or not.

Next, learners watched a movie during the “short film” module. The movie “Schwarzfahrer” (in English: fare dodger, Danquart, 1992) deals with xenophobia in everyday life. Afterwards, learners answered questions concerning their comprehension about the plot. The answers could be optionally discussed later in class. Finally, during the “online research” module, learners used the Internet to find definitions for several terms like culture or stereotype. For that purpose, an online browser was integrated into the program. The results were also discussed afterwards. Taken together, the learners’ own biography and the biography of foreigners played a major role within all parts of the program. All modules were designed to support understanding, experiencing, and facing foreignness in all facets of one’s own biography.

Design and procedure

A “pre and post” study design was implemented. Upon completion of the questionnaire and the IAT, the 61 students in the experimental group engaged in an instructional multimedia material that consisted of two teaching units that lasted for a total of 100 minutes. Finally, the post test was administered. The 18 students in the control group used the same assessments but did not participate in the multimedia learning.

Results

Due to the restrictions of the quasi-experimental design, we compared the pre-test values between the experimental group and control group. Descriptive values indicate that the sampling led to differences between both groups where the control group showed higher values in “Behavioral Control,” “Admission of Prejudice” and “Unprejudiced Self-perception”. Additionally, the control group showed lower values regarding “Xenophobia” and “Attitude Towards Foreigners” (see Table 1). A MANOVA showed that these difference were statistically significant ($F(5, 73) = 4.97; \ p < 0.001; \ \eta^2 = .25$). Due to these systematic pre-test differences we calculated the difference between pre- and post-test measures as subsequent dependant variables to avoid subsequent problems with this sampling bias. A MANOVA on these pre- and post-test differences showed no significant main effect, $F(5, 73) = 1.59; \ p = .17; \ \eta^2 = .10$. Additionally, there was no significant difference between the mean values of pre- and post-training sessions, $F(6, 55) = 0.35$, $p = .91, \ \eta^2 = .04$.

<table>
<thead>
<tr>
<th>Scale/Tests</th>
<th>Subscale</th>
<th>Experimental group</th>
<th>Control group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pretest $M$ (SD)</td>
<td>Posttest $M$ (SD)</td>
<td>Pretest $M$ (SD)</td>
</tr>
<tr>
<td>Motivation to unbiased</td>
<td>Behavioral control</td>
<td>2.05 (0.54)</td>
<td>2.05 (0.57)</td>
</tr>
<tr>
<td>behavior</td>
<td>Admission of prejudice</td>
<td>1.94 (0.60)</td>
<td>1.94 (0.57)</td>
</tr>
<tr>
<td></td>
<td>Unprejudiced Self-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>representation Xenophobia</td>
<td>2.90 (0.69)</td>
<td>2.98 (0.70)</td>
</tr>
<tr>
<td>Xenophobic Attitude</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>towards foreigners</td>
<td>Xenophobia</td>
<td>3.48 (0.71)</td>
<td>3.43 (0.93)</td>
</tr>
<tr>
<td>IAT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-212.89 (578.86)</td>
<td>-481.82 (713.7)</td>
<td>-356.57 (348.9)</td>
</tr>
</tbody>
</table>

Regarding the IAT pre-test scores, an ANOVA showed no significant group differences, $F(1, 70) = 1.59; p = .48; \ \eta^2 = .007$. However, the ANOVA comparing the post-test results showed that the experimental group had significantly lower scores on automated negative self-concept related to foreignness, $F(1, 70) = 3.07; p = .04; \ \eta^2 = .04$; one-sided. Subsequent t-tests for dependent samples showed a significant difference between pre- and
post-test implicit measures, \( t(61) = 4.23, p < .01, d = .41 \), for the experimental group. Descriptive values revealed that participants in the experimental group had a lower negative self-concept related to foreignness after the treatment than before. Regarding the control group, the difference between pre- and post-test was also statistically significant, \( t(17) = -1.95, p < .01, d = .42 \). In contrast to the experimental group, the control group participants had higher values in the post-test than in the pre-test, indicating that they had a higher negative self-concept related to foreignness.

**Discussion**

The study examined the consequences of intercultural instructional multimedia material on implicit and explicit attitudes towards foreigners. It was assumed that the program entitled, “Intercultural Games” would have a positive impact on the perception of others and would lead to a positive change of attitudes compared to a control group without intervention. Various explicit measurement methods were used such as, the “Motivation to Unbiased Behavior” scale by Banse and Gawronski (2003). An implicit association test was also used to measure the implicit xenophobic self-concept before and after the treatment.

The main question was to determine if attitudes can be changed after short didactical intervention sequences. Following the suggested General Xenophobic Model, we assumed that information processing regarding change in xenophobic attitudes could take an explicit or an implicit route, depending on the route of information preocessing.

Similar to previous research findings (e.g., Cochran-Smith, 2003; McDiarmid & Price, 1993; Wasonga, 2005), we were not able to find significant changes in explicit attitudes. However, there was a change at an implicit level related to the xenophobic self-concept of participants. While there was a significant decline in hostility toward strangers in the experimental group, there was an unexpected increase in the control group. The finding that there was a difference on an implicit level, but not an explicit level is not surprising because implicit and explicit measurement methods often do not show the same results, even when the same constructs have been measured (cf. Boysen & Vogel, 2008). One possible explanation is the disposition to social desirable answering that can influence the results, especially when considering ethical and moral issues. However, such a divergence was also measured in relation to other topics (cf. Greenwald & Nosek, 2001). Thus, moderator-variables need to be analysed for explaining such divergent outcomes. For example, Hofmann, Gawronski, Gschwendner, and Schmitt (2005) found evidence of a moderator effect between awareness and adjustment, thereby suggesting that adjustment effects may be more pronounced under conditions of high awareness. For Lambert, Payne, Ramsey, and Shaffer (2005), the perception of out-group homogeneity was a moderator for a high correlation between self-reported attitudes and reaction time based data.

Regarding aggressiveness, social desirability can be a moderator for a high coherence between implicit and explicit results. In relation to anxiety, desirability is not a moderator (Egloff & Schmukle, 2004). Thus, moderating variables may explain the divergence between implicit and explicit measures. However, in this study, moderators were not evaluated.

Based on the results of the study, it was shown that the IAT had a highly predictive power, especially regarding spontaneous behavior (cf. Gawronski & Conrey, 2004). Various studies confirm the link between spontaneous behavior and the results of the IAT (e.g., Asendorpf, Banse & Mücke, 2002; Egloff & Schmukle, 2004).

What does this mean in relation to this study? Due to the training with the instructional multimedia material, there was a positive change in spontaneous acts performed toward strangers. Thus, changing assessment procedures lead to a positive change of behavior toward foreigners. The procedures are part of the short-term processes of the GXM. If intercultural education are part of everyday instruction and not restricted to special subjects such as Geography or Philosophy, tolerant behavior and sustainable positive changes in attitudes can be achieved.

The results of this study are the first to support the validation of the GXM. The implicit measurement demonstrated the existence of the implicit way of information processing as a part of short-time processes related to the generation of xenophobic cognitions. Nevertheless, we were not able to confirm changes on the explicit route of information processing. Therefore, long-term interventions may be needed to replicate these findings.

Furthermore, it was shown that the range of impulsive and deliberate actions—based on evaluation and decision-making processes—could easily be influenced, even by didactical short-time interventions. Consequently,
repeated trainings could lead to long-term changes in attitudes. To verify prolonged changes, it will be necessary to conduct additional longitudinal studies. Limitations derive from the sample as used in this study. In this study, all participants were Caucasians. Although dark skin-color persons are a minority in the geographical location where the study was carried out, results of the evaluation of the material reveal these people seem to be judged more unfamiliar in daily life than they might in other geographical regions. Thus, the results are highly likely to differ in regions with a higher degree of different ethinical groups are living together.

**Broader implications**

The findings of this study might also have implications on current research on educational technology. While there is tradition of focusing on the cognitive and motivational aspects of learning with technology (cf. Astleitner & Hufnagl, 2003), emotional or attitudinal issues are rather rare. Instructional approaches such as the FEASP-approach (Astleitner, 2000), that are designed for emotionally sound instruction, are hardly implemented in educational practice or research. This study contributes to this area by integrating a cognitive and an attitudinal perspective in technology-based learning environments. This makes it necessary to change assessment strategies because traditional paper and pencil tests are not sensitive enough or are susceptible to social desirability. We attempted to overcome these known problems by using explicit and implicit measures, similar to how it has already been done in domains where technology might influence attitudes, such as assessing effects of violent computer games on aggression (Schrangl & Zumbach, 2012). This could also be an approach suitable for other domains, such as moral education or education for democracy, where technology-based learning environments could be implemented.

**References**


Danquart, P. (Director). (1992). Schwarzfahrer [Fare dodger] [Short film]. Germany.


The Effect of the Digital Classroom on Academic Success and Online Technologies Self-Efficacy

Mehmet Arif Ozerbas¹ and Bilge Has Erdogan²

¹Faculty of Educational Science, Gazi University, Ankara, Turkey // ²Ministry of National Education, Ankara, Turkey // ozerbas@gazi.edu.tr // bilgehas@windowslive.com

(Submitted July 13, 2015; Revised December 13, 2015; Accepted February 14, 2016)

ABSTRACT

This study aimed to observe whether the learning environment created by digital classroom technologies has any effect on the academic success and online technologies self-efficacy of 7th grade students. In this study, an experimental design with a pre-test/post-test control group was used. The research was conducted with 58 students in a secondary school of Ankara. At the beginning of the study, the students in the experimental and control groups used the “Academic success test” developed by the researcher and “Online Technologies Self-Efficacy Scale” developed by Miltiadou and Yu (2000) as pre-tests. At the end of the four-week experimental period, both groups used the same scales as post-tests. The results of this research have indicated that there is a meaningful difference in terms of academic success in favour of the experimental group. However, it has been shown that there is no meaningful difference in students’ online technologies self-efficacy. This research results are expected to provide useful information about the digital classroom’s being used at schools. In addition, when the importance digital classroom gave to interaction between parents and the school was thought, including parents into the research process may help to see the shortcomings clearly and develop the digital classroom.

Keywords

Digital classroom, ICT, Technological integration, Online technologies self-efficacy

Introduction

The rapid development of information and communications technologies (ICT) inevitably affects education, just as they affect everything related to people. Today, there are studies on the positive and negative effects of digital technologies that we are used to seeing in the classroom. In this context, each day there have been more studies on the integration of new digital tools with education (McGee & Diaz, 2005; Bullock, 2011; Gu, Zhu & Guo, 2013; Mishra & Koehler, 2006; Toledo, 2005; Tondeur, Van Keer, Van Braak, 2008; Vanderlinde & Van Braak, 2010; Wang & Woo, 2007). The common purpose of studies related to the integration of technology with education is to successfully integrate technology into the classroom environment. In fact, bringing technology into the classroom does not necessarily mean the integration of technology with education (Coklar, Kılıçer, & Odabaşı, 2007).

Learning environments enriched with technology have shifted from simple computer labs to highly technological environments equipped with computers, projection machines, internet connection and communications technology (Ott, 2000), and started to be named as digital classrooms with ICT opportunities. Any studies can examine the types of technology that digital classrooms have, are required to have and the technological competence level required by the teachers and partners of digital classrooms. What all of these possible studies foresee may be of great significance in terms of revealing the effects of the use of technology in the classroom environment.

The main purpose of integrating many digital technologies into the learning environment is increasing the quality and success of education. Looking through the related literature indicates that there are a number of studies that support using digital technologies in classrooms increases students’ academic success (Chen et al., 2013; López, 2010). According to Sezgin’s (2002) research, the classroom environment enriched with multimedia class software increases students’ academic success level. Similarly, Aktümen and Kacar (2003) have determined that using a computer and an internet connection in the classroom increases students’ success. On the other hand, some research shows that there are no positive effects of using technology in the classroom up students’ academic success (Dunleavy & Heinecke, 2008). Aktas, Alioğlu and Vardar (2007) determined that the academic success of students using information and communications technology is lower.

All these studies have shown that the use of various digital technologies in classroom environment may have positive or negative effects on academic success (Chen et al., 2013; Mashhadi & Kargoziari, 2011; Brown, 2011; López, 2010; Dunleavy & Heinecke, 2008; Rollins & Almeroth, 2004). At this point, it could be thought that...
efficiency in using these technologies is important for students in order that they can be successful both in and outside the classroom with digital technologies. Regarding that digital classrooms use online technologies both in and outside the classrooms, the level of students’ online technologies self-efficacy could be seen important in order to raise academic success. Online technologies self-efficacy defined by Miltiadou and Yu (2000) being suitable for the studies on digital classroom contains internet competencies, synchronous interaction and also asynchronous interaction. That is, the level of the students’ self-efficacy in these subtitles may affect academic success. Supporting this thought, Chang et al. (2014) have determined that there is meaningful relation between online college students’ internet self-efficacy and performances. Likewise, Tsai and Tsai (2003) have come to the decision that the students whose internet self-efficacy is higher can reach information more easily and learn better than the students whose self-efficacy is lower. This results in that online technologies self-efficacy seems of great importance for students in digital classrooms in terms of efficiently using the classroom equipment. There are research results showing that a student with a high level of online technologies self-efficacy is more successful in learning environments equipped with digital technologies (Miltiadou & Savenye, 2003; Wang, Shannon & Ross, 2013). Furthermore students who are more involved with digital technologies have a higher level of online technologies self-efficacy (Eastin & LaRose, 2000). Thus, the online technologies self-efficacy level of students is of great importance in terms of achieving the success expected from digital classrooms. Some research indicates that it is possible to increase the online technologies self-efficacy level through educational processes (Torkzadeh & Van Dyke, 2002). At this point, it is possible to study whether digital classes are effective on increasing the online technologies self-efficacy level.

In addition to the research mentioned above, the research conducted with a systematically formed digital classroom can be summarized as a general frame of the digital classroom (Beach, 2012; Garavaglia & Ferrari, 2012; Gordon, 2003; Mashhadi & Kargozari, 2011; Moran, 2006; Sultan, Woods & Koo, 2011), digital classroom equipment (Rollins & Almeroth, 2009) and student and teacher perception on the digital classroom (Garavaglia, Garzia & Petti, 2012; Roberts, 2007). Looking through the related literature, it can be seen that there are very few experimental studies on the digital classroom and the experimental processing made was limited to few digital classroom components such as student devices, network, smart board and etc. This research has been conducted in a learning environment that includes six components of the digital classroom which are student devices, teacher devices, communications network, class sharing screen, classroom servers and device management system (Liang et al., 2005) to fill this gap in the literature. This research, which focuses on the academic success and online technologies self-efficacy levels of students, is expected to make a good contribution to the literature.

Digital classroom

Today, information and communications technology (ICT), which enables fast access to information from anywhere, has been an indispensable part of our lives. ICT means any visual, audial, printed or written means that enables reaching and forming information (Cavas, Kılıç & Twining, 2004). Rapid development of ICT and the opportunities it provides have come up with the idea of using it in education. In this context, ICT has started to take its place in learning environments. These classrooms, in which ICT facilities are available, have started to be named digital classrooms. Roberts (2007) defined the digital classroom as an authentic environment equipped with digital technologies in which students have active participation and take responsibility for their own learning.

Digital classrooms can use virtual classroom technologies to provide learners with access to information, communication and participation in the learning process from anywhere (Postman, 1995). Unlike other classrooms, digital classrooms are equipped with ICT and other technologies. Students have computers, the Internet and other equipment in these classrooms. Devices used for applications can vary from mobile phones and personal digital assistants (PDA) to computers and electronic dictionaries (Liang et al., 2005). Moreover, an education site belonging to the classroom or school that is directed by the teacher is used for classroom activities and other activities on a regular basis.

Learning environment in digital classrooms

The digital classroom needs a well-planned classroom environment with specific features, rather than a random environment involving various technological devices. The classroom environment should have various components. The digital classroom environment may have many components. As seen in Figure 1, Liang et al. (2005) compiled components which digital classroom should have under six main titles as student devices,
teacher devices, communications network, class sharing screen, classroom servers and device management system.

- **Student Devices:** Personal mobile information devices for each student.
- **Teacher Devices:** Teacher’s personal computer or shared host computer.
- **Communications Network:** A communication element, including face-to-face peer communication, the wireless local area network or Internet connection.
- **Class Sharing Screen:** A sharing screen controlled by the teacher, which can be followed by a group or an individual.
- **Classroom Servers:** A system with an interface for open class application, which is formed of a learning management system and content management system, and enables connection with a comprehensive online learning community.
- **Device Management System:** The element that helps manage the teacher and student devices and other components (storage, charge, mobility).

![Digital classroom learning environment](image)

**Figure 1.** Digital classroom learning environment (Liang et al., 2005)

The best assistant for cooperatively carrying out the teaching-learning activities within a certain plan in the digital classroom is the common website belonging to the classroom. This website, in which each student can log in with their user name and password and the activities are prepared by the teacher in advance, is very significant in terms of coordinating the lesson. The site is a well-organized education environment prepared within certain criteria, rather than a random digital environment.

**Online technologies self-efficacy**

Self-efficacy means a person’s belief in his/her ability to succeed in specific situations (Bandura, 1977). Self-efficacy has effects on students’ cognitive, motive, emotional and selective processes. Upon the control of the incidents affecting people’s lives, no way of thinking is more important than their judgements towards their own capacities (Bandura, 1986). This is why self-efficacy is given importance in students’ learning. Compared to students who do not have confidence in their learning capacities, the students who believe in their high learning capacities study harder, show more willingness to participate in activities and have more efficient skills in coping with difficulties, thus they become more successful (Pajares & Shunk, 2002).

A student learning in digital classrooms needs to have self-efficacy so that he/she can reach information sources through digital tools and use these sources efficiently. In other words, online technologies self-efficacy is significant for students to achieve success in digital classrooms. Because students believe that their self-efficacy level determines their learning, motivation and academic success (Bandura, 1993).

According to Bandura (1977; 1986), self-efficacy is formed by four information sources. These are experience, vicarious experience, social persuasion and psychological factors. It can be said that these resources also develop online technologies self-efficacy (Miltiadou & Savenye, 2003). Regarding a student who learns in a learning environment equipped with digital technologies, it is an expected result that the student develops online
technologies self-efficacy. Regarding that self-efficacy may change depending on personal experiences, familiarity with the technologies used in any activities can be given as an example (Horzum & Çakır, 2009).

Recently, it has been popular to use digital technologies to reach the desired goals in educational activities. It is predicted that the digital classrooms, which involve many technological devices, will improve education’s success level. However, without a successful technological integration, it is unreasonable to claim that it is possible to reach these goals for students without the necessary online technologies self-efficacy. Due to these reasons, this research studies the effects of digital classrooms on academic success and online technologies self-efficacy. This study may provide decision-makers in education with some findings for the application about the use of digital classroom at schools. In addition, the information reached through the application may be presented to the parents who want their children to have education in digital classrooms (Leung, 2003). In this context, this research looks for answers to the following questions.

- Are there any differences between the academic success grades of students learning in digital classrooms (experimental group) and those who learn in a classroom without digital technologies (control group)?
- Are there any differences between the online technologies self-efficacy grades of students learning in digital classrooms (experimental group) and those who learn in a classroom without digital technologies (control group)?

**Methods**

**Design**

In this study, an experimental design with a pre-test/post-test control group was used to determine the effects of digital classrooms on academic success and online technologies self-efficacy. Two groups are determined through random selection for experimental design with the pre-test/post-test control group, and these groups are measured before and after the application and differences between the two measurements are evaluated (Büyüköztürk, 2010). The independent variable of this research in which the effect of the independent variable on the dependent variable is studied, is digital classroom application, and dependent variables are determined as students’ academic success and online technologies self-efficacy levels. Academic success test and online technologies self-efficacy perception scale were applied as pre-test with the Experiment and control groups before the process and these tests were applied at the end of the four-week experiment period as post-tests. The detailed information about these tests were given in instruments part.

**Working group**

*Features of the working group*

The research was conducted in the Gaziosmanpasa Secondary School, located in the province of Sincan, Ankara. Two 7th grade classrooms were chosen through random selection, and class A, which involves 32 students, was determined as the experimental group, and class B, with 31 students, was chosen as the control group. Five of the students in the control group who were absent for more than two weeks during the application process are not included in the statistical processes and analyses. Thus, the control group has 26 students.

The students in the experimental and control groups are given in Table 1 below.

| Table 1. Gender distribution of the students in the experimental and control groups |
|-------------------------------------|-----------------|-----------------|--------|
|                                     | Experimental group | Control group | Total  |
| Girls                               | 16               | 11              | 27     |
| Boys                                | 16               | 15              | 31     |
| Total                               | 32               | 26              | 58     |

**Teaching materials**

Circle and Circular region units from 7th grade Maths Curriculum were chosen for this work. Accordingly, 16 hours of application was completed in the form of 4 hours per week in 4 week-period. ICT classroom of the school was used as the digital classroom environment in which Experiment group was taught Maths. A computer for each student, a smart board for the classroom, a copier machine to get needed documents fast and a scanner
were supplied for the ICT classroom mentioned. Besides printed documents, a website was designed through which the students could join teaching activities also when they weren’t at school, communicate with each other, access the teacher when necessary logging in with their private password and username. There were also many tabs including some videos, pictures, homework, online exams, lesson programs and activities related to the topic on the classroom website. The system let the parents to be enrolled as knowledge users and get the online exam results of the students via e-mails. In this way, the students, the parents and the system manager, the teacher, were informed in each phase. In contrast to these, control group went on their activities with printed documents in a classroom with no digital Technologies.

**Instruments**

In this study, students used a 24-item academic success test, which was developed by the researcher to determine their math success levels before and after its application. Besides, the “Online Technologies Self-Efficacy Scale” developed by Miltiadou and Yu (2000) was used to measure students’ online technologies self-efficacy levels.

**Academic success test**

To measure the students’ academic success levels, the researcher developed a success test to be applied as a pre-test and post-test regarding educational attainments. The content of this test includes features of circles, angles of circles, length of circles and area of circles. While preparing test items, the acquisitions for 7th grade math lessons instructional programme of Ministry of Education National Education have been taken as reference. All of the high school and university entrance exams were taken into consideration when preparing the test items. A table of specifications was prepared in light of acquisitions to determine the test items. A multiple-choice test of 24 items was prepared according to expert opinions to measure the desired acquisitions. A pre-pilot application was held to assess the validity and reliability of the test items. After the application, it was seen that the difficulty index varied from 0.00 to 0.86 and discrimination index varied from 0.00 to 0.68. High rate of item discrimination raised test validity. If the discrimination index of the items is 0.40 or over, the item becomes quiet good; if it is between 0.30 – 0.39, the item is seen as good. If it is between 0.20 – 0.29, it may be used just when it is seen compulsory but still it needs some revision and improvement on it. If it is 0.19 or lower, the item is really weak. If it cannot be improved with some corrections, it should be removed out of the test (Tekin, 2000).

After pre-pilot application item number 16, which has a distinctiveness value lower than .30, has been changed by taking the review of the expert into consideration. No more change was made in any other items but just the mentioned ones. Finally one hundred twenty four students took the test to measure the validity and reliability of the test items. As a result of the application, Alpha was measured as .85, average item difficulty was .58 and item discrimination index of the items is 0.

**Table 2. CFA Results related to online technologies self-efficacy scale**

<table>
<thead>
<tr>
<th></th>
<th>( \chi^2 )</th>
<th>df</th>
<th>( \chi^2/df )</th>
<th>RMSEA</th>
<th>CFI</th>
<th>GFI</th>
<th>AGFI</th>
<th>RMR</th>
<th>NFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTOO</td>
<td>438.5</td>
<td>202</td>
<td>2.17</td>
<td>0.077</td>
<td>.98</td>
<td>.83</td>
<td>.79</td>
<td>0.052</td>
<td>0.96</td>
</tr>
</tbody>
</table>

According to the results of the exploratory factor analysis (EFA), the original OTSES is a one-factor scale of 29 items, which is prepared in four sub-scales as Internet competencies, Synchronous interaction, Asynchronous interaction I and Asynchronous interaction II. Four new items have been added to the original scale according to expert opinions. Thus, according to the EFA results of this 33-item scale, all scale items are assembled in one single factor. The factor load of this single factor scale, in which the Cronbach’s Alpha coefficient was calculated as .96 and 46.6 % of total variance was imparted, varied between .53 and .77. As a result of the validity and reliability tests of the Online Technologies Self-Efficacy Scale, the KMO (Kaiser-Mayer-Olkin) value of the scale was calculated as .94 and Barlett’s \( \chi^2 \) coefficient was 453.5 \( (p < .001) \). These values can be interpreted as the scale is appropriate for Exploratory Factor Analysis (EFA).
Upon the confirmatory factor analysis (CFA) of the OTSES, as items number 10, 12, 16, 18, 20, 22, 23, 24, 27, 32 and 33 were found out to be the items leading chi-square decrease in the model, they were removed from the scale. The criteria for fit indices included $\chi^2$/df being less than 5, GFI being more than 0.90, CFI being more than 0.95 and RMSEA being 0.08 or less (Brown, 2015). According to Table 2, one-factor structure of OTSES was verified by CFA and revealed that this one-factor model had an acceptable level of goodness of fit index ($\chi^2$/df = 2.17, RMSEA = 0.077, CFI = 0.98, GFI = 0.83).

Data analysis

The SPSS 20 package program has been used for data analysis, and two-factor ANOVA (repeated measures) has been used for repeated measurements applied to the experimental and control groups.

Results

The Result related to academic success

The two-factor ANOVA results are given in Table 3 below, showing whether there is a meaningful difference between academic success rates of the students in the experimental and control groups who were exposed to different educational processes before and after the experiment.

<table>
<thead>
<tr>
<th>Source of variance</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inter-subjects</td>
<td>18306.393</td>
<td>57</td>
<td>1140.532</td>
<td>3.721</td>
<td>.059</td>
</tr>
<tr>
<td>Group (Experimental/Control)</td>
<td>1140.532</td>
<td>1</td>
<td>1140.532</td>
<td>3.721</td>
<td>.059</td>
</tr>
<tr>
<td>Error</td>
<td>17165.861</td>
<td>56</td>
<td>306.533</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intra-subjects</td>
<td>11566.634</td>
<td>58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measurement (pre/post)</td>
<td>4280.355</td>
<td>1</td>
<td>4280.355</td>
<td>40.558</td>
<td>.000</td>
</tr>
<tr>
<td>Group*Measurement</td>
<td>1376.260</td>
<td>1</td>
<td>1376.260</td>
<td>13.041</td>
<td>.001</td>
</tr>
<tr>
<td>Error</td>
<td>5910.019</td>
<td>56</td>
<td>105.536</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>29873.027</td>
<td>115</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Analysing the Table 3, showing pre-test and post-test ANOVA results for the academic success test, it is seen that there is not a statistically meaningful difference between the pre-test and post-test academic success grades of the students in the experimental and control groups before and after the experiment [$F(1, 56) = 3.721, p > .05$]. This finding shows that the academic success levels of the students in the experimental and control groups do not differentiate, regardless of measurement differentiation (before and after the experiment).

According to Table 3, academic success grades of the students in the experimental and control groups (regardless of groups) vary depending on the classroom in which they learn [$F(1, 56) = 40.558, p < .05$]. In other words, applications in both the experimental and control groups increased students’ academic success.

When we look at Table 3, it is seen that the changes in students’ academic success in the experimental and control groups indicate a meaningful difference in favour of the experimental group [$F(1, 56) = 13.041, p < .05$]. That means the academic success level of the students in the experimental group, who learn mathematics in a digital classroom, is higher than the academic success of students in the control group, who learn in the classroom without any digital technologies. This effect can be due to using the digital classroom. Accordingly, it can be interpreted that using digital classrooms is a significant factor in increasing students’ academic success.

The Result related to online technology self-efficacy

The two-factor ANOVA results are given in Table 4 below showing whether there is a meaningful difference between online technologies self-efficacy levels of the students in the experimental and control groups before and after the experiment, depending on different applications.

During the measurements of the students in the experimental group and the students in the control group before and after the experiment it is observed that there is not a significant difference between the online technologies.
self-efficacy levels given in Table 4 \( F(1, 56) = .277, p > .05 \). This finding means that being in the experimental group does not statistically affect the level of online technologies self-efficacy perception.

<table>
<thead>
<tr>
<th>Source of variance</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inter-subjects</td>
<td>23.647</td>
<td>57</td>
<td>.116</td>
<td>.277</td>
<td>.601</td>
</tr>
<tr>
<td>Group (Experimental/Control)</td>
<td>.116</td>
<td>1</td>
<td>.116</td>
<td>.277</td>
<td>.601</td>
</tr>
<tr>
<td>Error</td>
<td>23.531</td>
<td>56</td>
<td>.420</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intra-subjects</td>
<td>8.889</td>
<td>58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measurement (pre/post)</td>
<td>.140</td>
<td>1</td>
<td>.140</td>
<td>.919</td>
<td>.342</td>
</tr>
<tr>
<td>Group x Measurement</td>
<td>.236</td>
<td>1</td>
<td>.236</td>
<td>1.554</td>
<td>.218</td>
</tr>
<tr>
<td>Error</td>
<td>8.513</td>
<td>56</td>
<td>.152</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>32.536</td>
<td>115</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

According to Table 4, online technologies self-efficacy perception grades of the students in the experimental and control groups (regardless of groups) do not vary depending on the classroom in which they learn \( F(1, 56) = .919, p > .05 \). In other words, the applications have not been effective in increasing students’ self-efficacy levels.

Analysing Table 4, it is seen that the changes in the online technologies self-efficacy perception level of the students in the experimental and control groups are not different. \( F(1, 56) = 1.554, p > .05 \). According to this finding, after application, the online technologies self-efficacy level of the students in the experimental group is not different from the self-efficacy level of the students in the control group. This finding means that the digital classroom is not effective in improving students’ online technologies self-efficacy levels. The reason for this finding can be the fact that the students live in a city, where it is easy to access technology, and almost all of the students have computers.

Conclusion

This research, which studies the effects of digital classrooms, is looking for answers to the question, “Does the digital classroom have any effects on academic success and online technologies self-efficacy?” To find the answers to this question, two classrooms were prepared: a digital classroom for the experimental group and an ordinary classroom without any digital technologies for the control group. The research’s results indicated that the academic success of the students in the digital classroom is meaningfully higher than the students in the classroom without any digital technologies. However, no meaningful differences have been found between the online technologies self-efficacy levels of the two groups.

The research findings show that the academic success of the students in the experimental group, who learn in a digital classroom, is higher than the success of the students in the control group, who learn in the classroom without any digital technologies. One reason for this result can be the concretization of abstract math subjects through the use of computer software in courses. It can also be said that doing digital-based activities, which are visually and auditory more attractive, increases students’ motivation, and thus resulted in higher success. Moreover, another reason can be the website used in the digital classroom, which enables students to access classroom documents and review their courses and also provides for communication among classroom partners. Looking through the literature will show that there are several other studies parallel to this research resulting in positive effects of digital classrooms on academic success (Aktümen, & Kacar, 2003; Aloraini, 2012; Chen et al., 2013; Gulek & Demirtas, 2005; Güven, & Sülün, 2012; Huppert, Lomask & Lazarowitz, 2002; López, 2010; Sezgin, 2002; Weathersbee, 2008). The point which many researchers in this field agree on is the belief that a technology well-integrated with the curriculum can have positive effects on students’ performance and academic success levels (Brown, 2011; Judge, 2005; Sinclair, 2009).

According to the results of this research, there is not a meaningful difference between the online technologies self-efficacy levels of the students who learn in the digital classroom and the students who learn in the classroom without digital technologies. The online technologies self-efficacy level of the students in both groups seemed to be very high, according to the pre-test results just before the application of this research. The reason for this finding can be the fact that the students live in a city, where it is easy to access technology, and almost all of the students have computers. The research results reaching in the result of the current researches, that online technologies self-efficacy of students is high, were released supporting the results of this study without having any applications (Puzziferro, 2008; Aesaert & Van Braak, 2014). In the formation of these results, the increase of
the students’ experience along with the results of the high speed in the development of technology and its being spread out may have been effective. Thus, experience may be the important determiner of self-efficacy (Isman & Celikli, 2009).

Only the use of digital equipments in the classroom does not turn the classroom into a digital classroom. When the digital classroom was taken as the realization of the planned instructional activities with the help of digital technologies whether at or outside the school, it is hoped that this work will present itself as a good model for the ones who are to work in this field. Also, this work carried out experimentally may present detailed information for the teachers who want to teach in digital classrooms.

The digital classroom is a very new research field. Because digital technologies are rapidly spreading in classroom environments, conducting more research on this subject will ensure the appropriate use of technology and achieve academic success. This research has been limited to maths lesson within the four-week application with 7th grade students. For researchers who will study this subject, it will be helpful to conduct qualitative research for longer periods without being limited to math courses to gather more detailed data. In addition, when the importance digital classroom gave to interaction between parents and the school was thought, including parents and the management staff of the school into the research process may help to see the shortcomings clearly and develop the digital classroom.

References


Aloraini, S. (2012). The Impact of using multimedia on students’ academic achievement in the college of education at King Saud University. *Journal of King Saud University*, 24, 75-82.


The Effectiveness Evaluation among Different Player-Matching Mechanisms in a Multi-Player Quiz Game

Fu-Hsing Tsai
Teacher Education Center, National Chiayi University, Taiwan // fhtsai@mail.ncyu.edu.tw

(Submitted July 30, 2015; Revised December 22, 2015; Accepted March 13, 2016)

ABSTRACT
This study aims to investigate whether different player-matching mechanisms in educational multi-player online games (MOGs) can affect students’ learning performance, enjoyment perception and gaming behaviors. Based on the multi-player quiz game, TRIS-Q, developed by Tsai, Tsai and Lin (2015) using a free player-matching (FPM) mechanism, the same game, but incorporating an automatic player-matching (APM) mechanism and Elo rating system, was developed in order to compare its effectiveness with the original TRIS-Q. The research findings indicate that students using this new player-matching mechanism of TRIS-Q acquired more knowledge of energy, experienced more enjoyment and exhibited more favorable gaming behaviors than did the students using the original TRIS-Q with FPM mechanism. Besides, after comparing the other TRIS-Q with APM mechanism, but using an automatic random player-matching mechanism, this study found that using Elo scoring for automatic matching was the more favorable player-matching mechanism in educational MOGs.

Keywords
Player matching, Educational multi-player online game, Effectiveness evaluation, Game-based learning

Introduction
Competition is ubiquitous in daily life. Political campaigns, sports events, business rivalries and academic contests hosted at schools are all competition-related (Czaja & Cummings, 2009). Competition has been defined as a social process that occurs when people are rewarded on the basis of how their performance compares with that of others performing the same task or participating in the same event (Coakley, 1997). Therefore competitive activities are typically associated with the ambition to win and outperform opponents (Kilduff, Elfenbein, & Staw, 2010). To win, participants draw upon their potential to perform more exceptionally (Tauer & Harackiewicz, 2004). For example, bicycle racers perform better when competing with other racers in a competition than when practicing alone (Triplett, 1898). Competition also triggers the intrinsic motivation of participants and the intention to continue to participate in competitive activities (Deci & Ryan, 1985; Harackiewicz, Barron, & Elliot, 1998). Additionally, competition enables people to learn from their failures, stimulates the learning motivation of participants and prompts the losers to improve themselves and continually strive toward victory (Lam, Yim, Law, & Cheung, 2004). In other words, competition exerts positive effects on performance improvement, participation motivation and learning motivation.

Although competition may enable losers to understand their flaws, and motivate them to learn, it may also reduce their self-confidence and learning motivation due to continual defeats (Fülöp, 2009). Thus, any forms of competition, whether direct, indirect, or cooperative, can also exert negative effects among the losers (Graham, 1976). However, according to Csikszentmihalyi’s flow theory, the flow state, which causes people to thoroughly immerse themselves in the activities, leading to optimal experience (Csikszentmihalyi & LeFevre, 1989), occurs when people concentrate on a competitive activity, and only when people’s skills and the difficulty of competition achieve a certain level of balance (Csikszentmihalyi, 1975; Rheinberg, 2008). In other words, if we attempt to get people immersed in a competitive activity, the skill level among competitors must be equal; when the disparity between skill levels is too large, it prevents the competitors from entering a state of flow, and the competitor with inferior skills can lose interest and confidence in the activities because of continual defeats and frustration. Hence, the ideal approach to reduce the negative effects of competitions is to create a fair competitive environment in which the skills of competitors are balanced, so everyone has a chance to win; that is, a fair player-matching mechanism is critical in competitions.

Because all games involve competition (Bright & Harvey, 1984; Crookall, Oxford, & Saunders, 1987), multi-player online games (MOGs) are also a typical competitive activity that includes multi-player competition against players. Also, MOGs include the general characteristic of digital games, such as fantasy, mystery and control, so they can elicit enjoyment from participants; thus, they have currently become an important recreation activity among young people. In other words, MOGs also reflect the positive effects which competition exerts, and attract many students’ involvement. Therefore, in recent decades, since Prensky (2001) began to advocate...
the use of digital game-based learning (DGBL), more and more researchers have studied educational MOGs; they integrate instructional contents and the features of MOGs, to promote student learning motivation and performance with numerous positive outcomes (Cheng, Kuo, Lou, & Shih, 2012; Tsai, Tsai, & Lin, 2015; Tsai, Yu, & Hsiao, 2012). All of these studies expected educational MOGs to achieve the positive effects of competition; that is, students can repeatedly improve their knowledge while engaging in games, in order to win, or outperform other players.

For example, Tsai et al. (2015) developed an educational MOG, TRIS-Q, in an online learning environment to promote students’ self-assessment and learning about energy by playing the MOG TRIS-Q is a multi-player quiz game which combines the tic-tac-toe game with multiple choice tests; that is, the first player to create a row of three pieces in a nine-square grid game board wins. Players need to answer random multiple choice questions related to energy knowledge at each step, when placing their piece on the game board (the game will be introduced in detail later). The game was developed so that students can continually and repeatedly improve their energy-related knowledge and participate in the game in order to defeat their competitors. The results of that study revealed that this game enhanced learning motivation and performance; in addition, providing immediate elaborated feedback resulted in optimal student learning. However, when the researchers also noted that the learning performance of the students playing the game might be affected by inappropriate player matching, they recommended further investigation of this problem. In other words, Tsai et al. (2015) also support the view that appropriate player matching is crucial in competitions.

Numerous scholars (Graepel & Herbrich, 2006; Véron, Marin, & Monnet, 2014) also maintain that providing a fair player-matching mechanism is critical in commercial MOGs. However, few studies have investigated this issue in educational MOGs. Therefore, the research suggestion by Tsai et al. (2015) is worth investigating. This study aims to follow their suggestion and research whether providing different player-matching mechanisms in TRIS-Q can positively affect learning effectiveness. To investigate this issue, the player-matching mechanism used in the TRIS-Q was first analyzed. This study found that TRIS-Q applied a lobby-based matching mechanism incorporated in past commercial MOGs; a game lobby is provided in which players can freely configure their own match (Berman & Bruckman, 2001). In this free player matching (FPM) mechanism, gamers can play with their friends or select opponents whose skills match their own. However, players may need to invest much time to search for opponents (Aikawa, Hei, Ogishi, Niida, & Hasegawa, 2013). Moreover, some players can repeatedly compete with the same opponents, resulting in cheating, whereby players gain points by conspiring to take turns beating each other.

Hence, many commercial MOGs currently adopt an automatic player-matching (APM) mechanism according to player skill, in order to provide a fair and effective competitive environment; these games always have a suitable skill rating system (Graepel & Herbrich, 2006). For example, a current popular commercial MOG, League of Legends, uses an APM mechanism and matches players who have nearly equal skill levels, primarily by using the Elo rating system (Shores et al., 2014) developed by Arpad Elo for assessing the relative skill levels of chess players, based on statistics (Elo, 1978). The traditional Elo rating system displays the score of each chess player from 0 to 3,000, based on the assumption that the performance of each player is a normally distributed random variable (Hacker & Von Ahn, 2009), so that the average player’s score is 1500. Because the initial skill level of each player is unknown, each player starts with a score of 1,500, which is then gradually modified according to the player’s performance in subsequent games (Glickman, 1995). The higher the score becomes, the higher the skill level a player possesses. Therefore, when two contestants compete, the competitor with the higher score is predicted to have the higher probability of winning the game. Assume \(R_A\) and \(R_B\) represent the rating scores of Contestants A and B, respectively. According to the Elo rating theory, the expected score of Contestant A winning the game can be expressed in the following formula (Elo, 1978):

\[
E_A = \frac{1}{1 + 10^{(R_B - R_A)/400}}
\]

Assume the original scores of Contestants A and B are 1,700 and 1,500, respectively. According to the aforementioned formula, the expected score of Contestant A (Contestant B) is .76 (.24); in other words, the higher the rating score, the greater the predicted probability that a player will win. Because the real skill level of each contestant is an unknown variable, the Elo rating system gradually adjusts the value representing the skill level of each contestant according to the following formula and the aforementioned expected scores:

\[
R'_A = R_A + K(S_A - E_A)
\]

where \(R'_A\) represents the new rating score of Contestant A after a game; \(R_A\) represents the original rating score of Contestant A. If Contestant A wins the game, then \(S_A = 1.0\); if the game is a draw, then \(S_A = 0.5\). If Contestant A
loses the game, then \( S_A = 0 \). Coefficient \( K \) is a floating coefficient that typically decreases as the rating score of a contestant increases. According to the aforementioned formula, if \( K = 30 \) and Contestant A wins the game, then \( R'_A = 1707.2 \); if Contestant A loses, then \( R'_A = 1677.2 \). Specifically, if Contestant A defeats Contestant B, then the score of Contestant A only increases by 7.2; if Contestant A is defeated by Contestant B, then Contestant A loses 22.8 points. This indicates that if the score of Contestant B is lower than that of Contestant A initially, then Contestant A gains fewer points by winning, but loses more points by losing. Thus, the score of each contestant is gradually adjusted according to this rating system from game-to-game. Therefore, based on this rating system, regardless of whether many contestants are in a competition, a contestant is not required to compete with each contestant successively. When such score calculation is implemented, the ranking order of all contestants is gradually adjusted, and a relatively stable rating score representing the skill level of each contestant is determined. Thus, the Elo rating system has been broadly applied in MOGs for APM mechanisms. The advantage of APM is that players can rapidly compete with random opponents, while the disadvantage is that players cannot choose their opponents by personal preference. For example, Claypool, Decelle, Hall and O’Donnell (2015) found that League of Legends players think that most games lack balance and are not fun, although an analysis of game data showed that most games are balanced.

In summary, competitive activities have both advantages and disadvantages. To enable people to immerse themselves in competitive activities and minimize the negative effects of such activities, creating a fair competitive environment is necessary. Therefore, appropriate player-matching mechanisms are crucial in MOGs. Currently, two types of player-matching mechanisms are generally incorporated in commercial MOGs, namely the FPM and APM mechanisms, both of which have strengths and weaknesses. However, few studies have examined which of these two types of mechanisms is more appropriate for application in an educational MOG. Therefore, in this study, the effectiveness of the FPM and APM mechanisms was investigated in depth. On the basis of Tsai et al.’s (2015) TRIS-Q, which is a multi-player quiz game based on the FPM mechanism (hereafter referred to as TRIS-Q-FPM), this study developed the same game, but incorporated an APM mechanism (hereafter referred to as TRIS-Q-APM), and subsequently compared the TRIS-Q-APM and TRIS-Q-FPM. In addition, according to the approaches employed in commercial MOGs, the Elo rating system was incorporated into TRIS-Q-APM as a game scoring mechanism and the basis for automatic player matching. However, to verify the applicability of the Elo rating system in the APM mechanism, two types of TRIS-Q-APM, one game involving APM based on Elo rating scores (TRIS-Q-APM-Elo) and the other game incorporating APM based on random matching (TRIS-Q-APM-Ran), were designed for this study. The TRIS-Q-APM-Ran game was used as an experimental control group involving an unfair player-matching mechanism. Specifically, the objectives of this study were as follows:

- Compare the effectiveness of the TRIS-Q-FPM and TRIS-Q-APM on the knowledge acquisition of students playing the games.
- Compare the degree of enjoyment among students who played the TRIS-Q-FPM game with those who played the TRIS-Q-APM game.
- Compare the gaming behavior of students who played the TRIS-Q-FPM game with those who played the TRIS-Q-APM game.
- Examine the applicability of the Elo rating system in the TRIS-Q-APM game.

**Methods**

**Research design, participants and procedures**

A non-equivalent pre-test–post-test control group design was adopted in this study to investigate how the multi-player quiz games with different player-matching mechanisms influence the learning effectiveness. Hence, the independent variable was the player-matching mode divided into three groups: one group that played the TRIS-Q-FPM (FPM group) and the other two groups that played the TRIS-Q-APM-Elo (APM-Elo group) and TRIS-Q-APM-Ran (APM-Ran group) games, respectively. The dependent variables were students’ acquisition of energy knowledge, enjoyment perception and gaming behaviors. Knowledge acquisition was measured based on the score that each participant acquired from the energy knowledge test. The enjoyment perception was assessed based on the score that each participant obtained on the game enjoyment scale. The gaming behaviors were measured based on each participant’s gaming records, including the number of games participated in, the number of questions answered in the game and the correct answer ratio in the game. In addition, the applicability of the Elo rating system in the TRIS-Q-APM game was explored. Thus, the participants’ Elo game scores and tic-tac-toe skills between two APM groups were specifically analyzed. The tic-tac-toe skill was assessed based on the score that each participant obtained from the tic-tac-toe ability test.
The study participants comprised students from three randomly sampled 12th grade classes (Classes A, B and C with 40, 41 and 33 students, respectively) at a high school in Kaohsiung, Taiwan. After excluding some participants who could not complete the experiment, the FPM group comprised 28 students from Class A (11 boys and 17 girls), the APM-Elo group included 27 students from Class B (9 boys and 18 girls) and the APM-Ran group consisted of 25 students from Class C (17 boys and 8 girls). In correspondence with the energy technology unit of the living technology course in these three classes, the experiment lasted a total of 4 weeks (one 50-min class per week). Before the experiment, a pre-test was conducted on student knowledge of energy. At the beginning of the experiment, the teachers instructed the participants on how to log into the online learning systems and play the multi-player quiz game. Subsequently, because all of the participants had acquired basic knowledge related to energy in a junior high school course before the experiment, the participants were asked to read the online learning materials constructed by Tsai et al. (2015), including three energy knowledge topics (sources of energy, application of energy, and energy conservation and new energy) for at least 20 min. After that, all of the participants began playing the game assigned to them. After the experiment ended, all of the participants were administered an energy knowledge post-test and tic-tac-toe ability test, and participants rated their perceived enjoyment when playing the game by using a scale.

**Instruments and materials**

**TRIS-Q-FPM**

To examine the effectiveness of the game-based assessment in online learning, Tsai et al. (2015) developed a TRIS-Q game. This game, referred to as TRIS-Q-FPM in this study, was administered to the FPM group in this study, and included in Tsai et al.’s (2015) online learning system. After logging into the online learning system and entering the game via a general web browser, a game lobby is displayed (Figure 1). This interface enables participants to freely search for opponents. Participants can investigate the performance (number of wins and losses) of online participants by clicking the user name in the user list, to interact with other online players publicly or privately, and to challenge someone to compete by typing in text. Subsequently, the participants can create a “game room” with or without password and wait for other participants to join, or join a game room created by other participants in the competing list, and begin playing.

![Figure 1. The interface of the game lobby in TRIS-Q-FPM](image)

When a game starts, a tic-tac-toe board is displayed on the computer screen (step 1 of Figure 2) and used by two competing players. The basic game rule is similar to that of a conventional tic-tac-toe game, in which the first player to create a row of three pieces wins. However, each time a player places a piece on the board, a random multiple-choice question appears selected from the database, which had 149 test items related to the three primary energy knowledge topics from the online learning contents in the online learning system (step 2 of Figure 2). The player must answer the question correctly to retain the piece on the desired board location (step 3-2 of Figure 2); if the player answers incorrectly, the opponent’s piece is placed at the location instead (step 3-1 of Figure 2). Also, when completing a turn and waiting for their opponent’s turn, the left side of the game screen will show the previous question, immediate feedback and elaborated feedback to show whether their answer is correct, and to offer clues relevant to the question for the player (step 3-1, 3-2 of Figure 2). Additionally, as depicted in Figure 1, the function of the ranking board and answer history are provided in the TRIS-Q-FPM interface. The ranking board lists the top 10 players with the most favorable performances in the tic-tac-toe game and the most correct answer ratios; the answer history function enables participants to review all of the test questions and clues they have answered.
Hence, if the participants want to be the top 10 players, they must try to repeatedly defeat their competitors for acquiring the game scores. At the same time, in order to defeat their competitors, players need to try to answer the energy questions correctly and constantly improve their energy knowledge in multiple ways. For example, the participants can improve their energy knowledge by reading each question’s clue or by rote when competing or using the answer history function. Students can also acquire energy knowledge by exiting the game for reading the online learning contents in the online learning system. Accordingly, this game can make learning happen.

TRIS-Q-APM

According to the TRIS-Q developed by Tsai et al. (2015), this study created an APM type of TRIS-Q, TRIS-Q-APM, for the APM groups. This game was designed to be embedded in Tsai et al.’s (2015) online learning system and can also be run in general web browsers. After entering the game, participants are shown the main game menu (Figure 3, left); they can click on the button at the lower right to begin the game.
applicability of the Elo rating system for player matching, the TRIS-Q-APM games were divided into two versions according to the automatic player-matching approaches: In the TRIS-Q-APM-Elo, the computer matches a player automatically with an opponent who possesses a similar Elo score to that of the player, and in the TRIS-Q-APM-Ran, a player is randomly matched with an opponent.

The game rules and the playing process of TRIS-Q-APM are the same as TRIS-Q-FPM: that is, during a game, a tic-tac-toe board is also displayed (Figure 4, left) and used by two competitors. The players are also required to answer questions (Figure 4, right) randomly selected from the same test items database with TRIS-Q-FPM, and feedback is generated according to the answers the players provide. In addition, as depicted in Figure 3, similar to the TRIS-Q-FPM, the TRIS-Q-APM features a ranking list and answer history function. In other words, the competition strategy and the multiple ways for making players obtain energy knowledge are also the same as for TRIS-Q-FPM. However, because the TRIS-Q-APM employs the Elo rating system to calculate game scores, the game performance ranking is based on the Elo scores of players rather than on the number of wins and losses.

![Figure 4. The screenshot of playing TRIS-Q-APM](image)

**Energy knowledge test**

A self-developed energy knowledge test was administered to understand the changes in the knowledge acquired by the students after they played the game. The test questions were created according to the content of the quiz game questions on energy knowledge. The test comprised 25 multiple-choice questions, and the full score was 100. The validity of the questions was verified by two high school teachers. According to a pre-test administered to 36 12th grade students, the Kuder–Richardson reliability of the test was .67, the average difficulty was .65 and the average discrimination was .31.

**Tic-tac-toe ability test**

A tic-tac-toe ability test revised by Tsai et al. (2015), according to the version created by Crowley and Siegler (1993), was administered to evaluate the tic-tac-toe skills of the participants. The test has 16 questions, each of which comprises an unfinished tic-tac-toe game; the participants select what they believe to be the optimal move required to finish the game. Each question has a correct answer regarding the optimal move. Thus, the tic-tac-toe skill of each participant is ascertained. The Kuder–Richardson reliability of the test was .78 through a pre-examination study with 36 12th grade students.

**Enjoyment perception scale**

A self-developed scale was used to evaluate the enjoyment the participants perceived when they played the multi-player quiz games. In reference to the scale on enjoyment developed by Downs and Sundar (2011), this scale was used to evaluate participant feelings on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The participants responded to the following eight questions: (1) I liked this game; (2) I found this game boring; (3) I enjoyed the competition while playing this game; (4) This game constantly frustrated me; (5) This game provided me with an enjoyable experience; (6) I think this was an exciting game experience; (7) I found this game challenging; and (8) I like this type of competitive game. The Cronbach α of this scale was an acceptable .85.
Results

Comparisons among different player-matching groups on students’ knowledge acquisition

To compare whether different player-matching mechanisms in a multi-player quiz game can affect students’ learning performance, the player-matching mechanisms (FPM, APM-Elo and APM-Ran) were designated as the independent variables, and the pre-test and post-test scores from the energy knowledge test were set as the covariates and dependent variables, respectively, for a one-way analysis of covariance (ANCOVA). Before the ANCOVA was performed, a test of within-class regression coefficient homogeneity was conducted; the result obtained was $F(2, 74) = .293, p = .747$, indicating that it fulfilled the basic assumption of ANCOVA. The result of ANCOVA was $F(2, 76) = 8.350, p = .001, \eta^2 = .180$, revealing that after the effect of the covariate was removed, the effect of different player-matching mechanisms on the energy knowledge post-test scores differed significantly. As revealed in the post hoc test (Table 1) using the LSD (least significant difference) method, the adjusted post-test average scores of the APM-Elo group (74.35) and the APM-Ran group (78.02) were both significantly higher than those of the FPM group (68.84). However, the average scores of the APM groups differed nonsignificantly. This finding revealed that the students who accepted the TRIS-Q with the APM mechanism acquired more energy knowledge than those who played the FPM mechanism game after the period of the experiment. It implies that the APM mechanism is more appropriate for application in TRIS-Q to promote students’ acquisition of energy knowledge.

Table 1. Post hoc comparisons (LSD) of mean differences on energy knowledge post-test

<table>
<thead>
<tr>
<th></th>
<th>FPM</th>
<th>APM-Elo</th>
<th>APM-Ran</th>
</tr>
</thead>
<tbody>
<tr>
<td>FPM</td>
<td>68.84</td>
<td>-</td>
<td>.015*</td>
</tr>
<tr>
<td>APM-Elo</td>
<td>74.35</td>
<td>-</td>
<td>.114</td>
</tr>
<tr>
<td>APM-Ran</td>
<td>78.02</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Note. $p < .05$.

Comparisons among different player-matching groups on students’ enjoyment perception

To compare whether different player-matching mechanisms in a multi-player quiz game can affect students’ enjoyment perception, the player-matching mechanisms (FPM, APM-Elo and APM-Ran) were designated as the independent variables, and the scores (average) from the game enjoyment scale were defined as the dependent variables for one-way analysis of variance (ANOVA). The result of Levene’s test for homogeneity of variances was $F(2, 77) = 2.872, p = .063$ and did not contradict the assumption of the homogeneity of variance. The ANOVA result indicated that the degrees of enjoyment experienced by participants in the three groups did not differ significantly, $F(2, 77) = 2.123, p = .127$. The LSD post hoc comparisons (Table 2) revealed that the average enjoyment score of the APM-Elo group (3.80) was significantly higher than that of the FPM group (3.47), but did not differ significantly from that of the APM-Ran group (3.58). This result revealed that most students had positive perception regarding the multi-player quiz game, no matter what player-matching mechanism was used. It means that game-based assessment can bring enjoyment for most students. However, the result also implies that using the Elo rating system for game scoring and automatic player matching is more appropriate for application in TRIS-Q to promote students experiencing enjoyment because the enjoyment perceived by the APM-Elo group was the highest among the three groups, and was significantly higher than that of the FPM group.

Table 2. Post hoc comparisons (LSD) of mean differences on enjoyment perception

<table>
<thead>
<tr>
<th></th>
<th>FPM</th>
<th>APM-Elo</th>
<th>APM-Ran</th>
</tr>
</thead>
<tbody>
<tr>
<td>FPM</td>
<td>3.47</td>
<td>-</td>
<td>.045*</td>
</tr>
<tr>
<td>APM-Elo</td>
<td>3.80</td>
<td>-</td>
<td>.202</td>
</tr>
<tr>
<td>APM-Ran</td>
<td>3.58</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Note. $^*p < .05$.

Comparisons among different player-matching groups on students’ gaming behaviors

To compare whether different player-matching mechanisms in a multi-player quiz game can affect students’ gaming behaviors, the player-matching mechanisms (FPM, APM-Elo and the APM-Ran) were designated as the independent variables, and the gaming behaviors observed in this study (e.g., the number of games participated
in, the number of questions answered in the game and the correct answer ratio in the game) were defined as the dependent variables for one-way ANOVA.

First, regarding the total number of games participated in, the result of Levene’s test before the ANOVA was $F(2, 77) = 2.817, p = .066$ and corresponded with the assumption of the homogeneity of variance. The result of ANOVA was $F(2, 77) = 9.380, p < .001$, indicating that the number of games participated in differed significantly among the three groups. The LSD post hoc comparisons (Table 3) revealed that the average numbers of games participated in by participants in the APM-Elo group (17.22) and APM-Ran group (18.96) were significantly higher than that in the FPM group (11.96); however, the number of games engaged in by participants in the two APM groups differed nonsignificantly. This result showed that the students who accepted multi-player quiz game with APM mechanism participated in more competitions than those who played the FPM mechanism game during the period of the experiment. It implies that the APM mechanism is more appropriate for application in TRIS-Q to enhance student opportunities or motivations involving game-based assessment.

**Table 3. Post hoc comparisons (LSD) of mean differences on the number of games participated in**

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>FPM</th>
<th>APM-Elo</th>
<th>APM-Ran</th>
</tr>
</thead>
<tbody>
<tr>
<td>FPM</td>
<td>11.96</td>
<td>-</td>
<td>.002*</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>APM-Elo</td>
<td>17.22</td>
<td>-</td>
<td>-</td>
<td>.314</td>
</tr>
<tr>
<td>APM-Ran</td>
<td>18.96</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. * $p < .05$.

Second, regarding the number of questions answered in the quiz game, the result of Levene’s test before the ANOVA was $F(2, 77) = 1.161, p = .319$, and corresponded with the assumption of the homogeneity of variance. The result of the ANOVA was $F(2, 77) = 8.922, p < .001$, indicating that the number of questions answered by participants in the three groups differed significantly. The LSD post hoc comparisons (Table 4) revealed that the average numbers of questions answered in the APM-Elo group (54.70) and APM-Ran group (62.80) were significantly higher than those answered in the FPM group (41.36); however, the number of questions answered in the two APM groups differed nonsignificantly. This finding revealed that the students who accepted the multi-player quiz game with the APM mechanism answered more questions than those who played the FPM mechanism game during the same duration. It also implies that the APM mechanism is more appropriate for application in TRIS-Q to enhance student opportunities or motivations in participating in online tests.

**Table 4. Post hoc comparisons (LSD) of mean differences on the number of questions answered in the game**

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>FPM</th>
<th>APM-Elo</th>
<th>APM-Ran</th>
</tr>
</thead>
<tbody>
<tr>
<td>FPM</td>
<td>41.36</td>
<td>-</td>
<td>.010*</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>APM-Elo</td>
<td>54.70</td>
<td>-</td>
<td>-</td>
<td>.123</td>
</tr>
<tr>
<td>APM-Ran</td>
<td>62.80</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. * $p < .05$.

Finally, regarding the correct answer ratio in the quiz game, the results of Levene’s test before the ANOVA was $F(2, 77) = 9.397, p < .001$ and did not support the assumption of the homogeneity of variance; therefore, a Brown–Forsythe F test was applied for analysis (Field, 2009). The result revealed that the correct answer ratio in the multi-player quiz game differed significantly among the participants in the FPM and APM groups ($Brown–Forsythe F(2, 47.38) = 13.070, p < .001$). The Games–Howell post hoc comparisons (Table 5) showed that the average ratio of correct answers in the APM-Elo group (.85) and the APM-Ran group (.84) were both significantly higher than those of the FPM group (.73), but the two APM groups differed nonsignificantly regarding this gaming behavior. This finding revealed that the students who accepted the multi-player quiz game with the APM mechanism answered more questions correctly than those who played the FPM mechanism game during the same duration. It implies that the APM mechanism is more appropriate for application in TRIS-Q to enhance students’ correct answer ratios in game-based assessment.

**Table 5. Post hoc comparisons (Games–Howell) of mean differences on the correct answer ratio in game**

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>FPM</th>
<th>APM-Elo</th>
<th>APM-Ran</th>
</tr>
</thead>
<tbody>
<tr>
<td>FPM</td>
<td>.73</td>
<td>-</td>
<td>.001*</td>
<td>.002*</td>
</tr>
<tr>
<td>APM-Elo</td>
<td>.85</td>
<td>-</td>
<td>-</td>
<td>.809</td>
</tr>
<tr>
<td>APM-Ran</td>
<td>.84</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. * $p < .05$. 220
Applicability of the Elo rating system in the multi-player competitive quiz game

To examine the applicability of the Elo rating for player matching in multi-player quiz games, an analytical comparison was performed specifically on the two APM groups; it incorporated the Elo rating system to calculate the game scores. In both TRIS-Q-APM-Elo and TRIS-Q-APM-Ran, each player started with a game score of 1,500, and the k value was designated as 30. According to the theory of the Elo rating system, the overall average game score of the participants should be 1,500 and the scores should be normally distributed. Therefore, the score statistics and distributions of the two APM groups were first compared. The average score of the APM-Ran group was 1,517.88 ($SD = 43.22$), and that of the APM-Elo group was 1,500.67 ($SD = 47.09$); thus, the average score of the APM-Elo group was closer to 1,500. Figure 5 presents the frequency and normal distribution graphs of the game scores obtained by the two APM groups, in which the scores of the APM-Elo and APM-Ran groups have a standard normal distribution and negative skewness, respectively. According to the average scores and score distributions of these two groups, the scores acquired by the players of TRIS-Q-APM-Elo correspond more closely to the theory of the Elo rating system.

Moreover, according to the theory of the Elo rating system, the Elo score obtained by each player from multiple games should finally reflect the player’s gaming skill. Because TRIS-Q involves board-gaming skills and energy knowledge, theoretically, a player with more advanced tic-tac-toe skills and energy knowledge should perform more outstandingly in TRIS-Q; in other words, the tic-tac-toe skill and energy knowledge of a player in TRIS-Q affects the gaming skill of the player. Therefore, if the Elo game score can reflect each player’s gaming skill, then each player’s tic-tac-toe skill and energy knowledge should affect their Elo game score. To determine which group’s game scores reflected the gaming skills of the participants more accurately, a multiple regression analysis was conducted. A simultaneous regression was executed on the two APM groups by using the tic-tac-toe ability test scores and the energy knowledge post-test scores as the independent variables, and the Elo scores of the participants as the dependent variables. The results (Table 6) revealed that in the APM-Ran group, the two independent variables had little effect on the Elo scores; the explanatory power of the regressions ($R$-squared) was only .040, and the regression effect was nonsignificant ($F(2, 22) = .453, p = .641$). Conversely, in the APM-Elo group, the two independent variables exerted a greater effect on the Elo scores; the explanatory power of the regressions ($R$-squared) was .257, and the regression effect achieved significance ($F(2, 24) = 4.156, p = .028$). In summary, compared to that of the APM-Ran group, the Elo game score of the APM-Elo group reflected the gaming skill of the participants more accurately, and corresponded more closely to the theory of the Elo rating system. It implies that using the Elo rating system for game scoring and automatic player matching is more appropriate for application in TRIS-Q to provide a fair competitive environment.

Table 6. Comparisons of regression analysis between two APM groups for the relationships among Elo score, tic-tac-toe ability and energy knowledge

<table>
<thead>
<tr>
<th></th>
<th>APM-Ran</th>
<th></th>
<th></th>
<th>APM-Elo</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Beta</td>
<td>t</td>
<td>p</td>
<td>Beta</td>
<td>t</td>
<td>p</td>
</tr>
<tr>
<td>tic-tac-toe ability</td>
<td>-.198</td>
<td>-.948</td>
<td>.353</td>
<td>.327</td>
<td>1.841</td>
<td>.078</td>
</tr>
<tr>
<td>energy knowledge</td>
<td>.016</td>
<td>.075</td>
<td>.941</td>
<td>.346</td>
<td>1.951</td>
<td>.063</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.040</td>
<td></td>
<td></td>
<td>.257</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F$</td>
<td>$F(2, 22) = .453, p = .641$</td>
<td></td>
<td></td>
<td>$F(2, 24) = 4.156, p = .028$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Discussion and conclusion

Based on a literature review, this study found that an appropriate player-matching mechanism is vital in educational MOGs, and also that the two types of player-matching mechanisms (FPM and APM) generally incorporated in commercial MOGs have their respective strengths and weaknesses. However, few studies have examined whether these two types of mechanisms in educational MOGs can affect students’ learning performance, enjoyment perception and gaming behaviors. Because the original TRIS-Q adopted an FPM mechanism for player matching, this study developed the same game, but incorporated an APM mechanism to compare its effectiveness with the original TRIS-Q. In addition, this study developed two types of APM mechanism for TRIS-Q: automatic player-matching by the Elo score and automatic random player-matching, to facilitate comparing the applicability of the Elo rating system in multi-player quiz games.

The experimental results indicated that students using an APM mechanism of TRIS-Q, whether based on Elo score or random matching, acquired more energy knowledge, experienced more enjoyment, invested more time playing the games and answering the questions, and answered more questions correctly than did the students using the original TRIS-Q with the FPM mechanism. The findings confirmed that the player-matching mechanism is important and can affect the learning effectiveness in educational MOGs. It was also consistent with the results of previous studies (Manweiler, Agarwal, Zhang, Roy Choudhury & Bahl, 2011), which had indicated that the FPM mechanism cost the participants excessive time in finding opponents, because compared with the TRIS-Q-APM, TRIS-Q-FPM severely reduced the number of games participated in and the number of questions answered. Besides, the most important finding is that it implied that the APM mechanism is more appropriate for application in TRIS-Q than is the FPM mechanism. Because this new player-matching mechanism accelerated the speed of finding the opponent in TRIS-Q, students using the APM mechanism of TRIS-Q perceived more enjoyment than did the students using the FPM mechanism of TRIS-Q. Accordingly, it further enhanced student motivation to spend more time playing the game. Because the more frequently that participants played the game, the more likely that the game helped students to improve their energy knowledge; therefore, this important finding seems to be reasonable.

However, according to the research findings, the players’ score distribution in TRIS-Q-APM-Elo was more consistent with the theory of the Elo rating system than was the players’ score distribution in TRIS-Q-APM-Ran, and the players’ Elo scores in TRIS-Q-APM-Elo represented their gaming skill more accurately. In other words, using the Elo scores to match players automatically was more meaningful than randomly matching opponents because this approach enabled a fairer competitive environment. This could be the reason why students using the TRIS-Q-APM-Elo experienced more enjoyment than those using TRIS-Q-APM-Ran. Hence, after comparing the two APM mechanisms of TRIS-Q, the finding of this study further implied that using Elo score matching for the APM mechanism is more appropriate for application in TRIS-Q than using the random matching for the APM mechanism. In summary, to create an effective player-matching mechanism for TRIS-Q, using APM mechanism and applying Elo scores as the standard for player matching constitute the best way. Because the TRIS-Q-APM-Elo not only provided a fair competitive environment by automatically matching players with similar Elo scores, but also created an effective player-matching mechanism for rapidly matching a new competition, the players’ enjoyment in TRIS-Q-APM-Elo was significantly greater than those who played the TRIS-Q-FPM. Consequently, students in TRIS-Q-APM-Elo were willing to spend more time playing the game repeatedly and thereby gaining more energy knowledge than those in the original TRIS-Q.

Although the research results in this study apparently can be applied widely in various educational MOGs, improvements upon this study in further studies are still needed. For example, the more accurate reasons why TRIS-Q-APM-Elo is the best-player-matching mechanism to promote higher learning performance need to be researched further in future studies. The major method by which students acquire energy knowledge from TRIS-Q-APM-Elo also needs to be explored in depth in the future. Moreover, the Elo ranking system using in TRIS-Q-APM-Elo can be studied and improved further. For example, the problem of freezing and inflation are well-known phenomena in the Elo rating system, especially when some players stop playing once they reach the top score or re-create multiple new accounts once they lose interest (Regan, Macieja & Haworth, 2012). Hence, in the future, larger samples are required to verify the results, and the duration of the experiment may be extended to expand the Elo score range of the participants, in addition to further examining the accuracy of the Elo scores and verifying if the problems of freezing or inflation still exist in TRIS-Q-APM-Elo. Alternatively, a dynamic k value in the Elo rating system is crucial to prevent the problem of inflation, and to accelerate the speed in finding players’ correct Elo value (Véron, Marin & Monnet, 2014). Therefore, the TRIS-Q-APM-Elo can be further enhanced by dynamically adjusting the k value for finding a suitable algorithm to rapidly converge players’ Elo scores and prevent the problems of the Elo rating system.
Acknowledgements

The work was supported by the Ministry of Science and Technology, Taiwan under contract numbers NSC 102-2511-S-415-013-MY2 and MOST 104-2511-S-415-010-MY2.

References


The Impact of Cultural Dimensions on Online Learning

Pilar Gómez-Rey¹*, Elena Barbera¹ and Francisco Fernández-Navarro²

¹Department of Psychology and Education, Universidad Oberta de Catalunya, Barcelona, Spain // ²Department of Quantitative Methods, Universidad Loyola Andalucía, Seville, Spain // pgomez_del_rey@uoc.edu // ebarbera@uoc.edu // i22fenaf@uco.es

*Corresponding author

(Submitted January 19, 2015; Revised January 20, 2016; Accepted January 31, 2016)

ABSTRACT

Due to the increasingly multicultural nature of e-learning environments, it is critical that instructors and instructional designers be aware of the importance of cultural factors in education and that they deliver culturally adaptive instruction. The main challenge of this paper is identifying the critical success factors for multicultural online learning from learners’ perceptions and their relationships with the six-dimensional Hofstede cultural model. Specifically, a categorization of the students’ perceptions is proposed by combining multiple correspondence and clustering analyses. To that end, two surveys were conducted in four e-learning universities in four different countries: Spain, the USA, China, and Mexico. Findings from clustering analysis show that learners are categorized according to their autonomy levels at the beginning of the educational process, and they are classified according to their satisfaction levels at the end of the instruction.

Keywords

Cross-cultural online learning, Hofstede cultural dimensions, Multiple correspondence analysis, Clustering

Introduction

Despite the relatively recent emergence of Internet-based training, online learning has evolved rapidly. Thus, blended learning (Bonk & Graham, 2012) and mobile learning (Baran, 2014) are some examples of the advances within this learning discipline. Different types of problems related to online learning have been analyzed in the literature, such as the study of the discipline’s effectiveness and efficiency (Cooper & Higgins, 2015) or the analysis of its advantages and disadvantages compared with other learning alternatives (Lim, Morris & Kupritz, 2014). In this research, we will focus our efforts on analyzing students’ perceptions with respect to online learning. In line with students’ perceptions in online education, several research papers have been published in recent years. For example, Paechter and Maier (2010) investigated students’ perceptions of online learning compared with traditional education, and Irwin, Ball, Desbrow and Leveritt (2012) analyzed learners’ perceptions about the use of Facebook as a tool to promote active learning.

Multicultural contexts are currently widespread in society. Although cultural diversity leads to individual and collective enrichment, several confrontations and misunderstandings between people can result from cultural differences, which also affect the education discipline. Culture greatly influences social behavior, communication, cognitive processes, and pedagogical technologies (Vatrapi, 2008). All of these elements are key components in the online learning discipline. Hence, the design of online programs requires cautious study of how people learn, what people learn and what people perceive as important to learn, depending on their culture. Understanding the set of cultural and learning/teaching features will help the educational community to provide better quality yet also culturally sensitive instruction.

Different authors have researched how culture affects societies through years. For example, Trompenaars and Hampden-Turner (1998) identified seven cultural dimensions in order to determine how people solve problems and reconcile dilemmas. Schwartz (1999) presented a theory of basic human values based on seven cultural dimensions. Inglehart (2008) categorized countries according to two values: (i) survival values versus self-expression values and (ii) traditional values versus secular-rational values. Finally, Hofstede (2011) tried to clarify the values that prevail in different national cultures, first, through a four-dimensional (4D) model, second, through a five-dimensional (5D) model and, third, through a six-dimensional (6D) model.

The specific objectives of this research are as follows: (1) To identify the most important variables in the online learning process, as perceived by students with significantly different cultural backgrounds and (2) to link our findings to the Hofstede 6D model. Thus, the main differences of this inquiry with respect to the online learning literature that analyzes these problems are as follows:
We analyzed the main educational variables that define a complete educational process (including the stage where the students are not yet enrolled in the course, the variables involved while the student is taking the course, and finally the variables implicated after the course), unlike state-of-the-art research works that focus on specific parts of the online course. For example, Liu, Liu, Lee and Magjuka (2010) focused their attention on how culture affects course design variables (e.g., language, communication tools, and time zone differences); Marambe, Vermunt and Boshuizen (2012) analyzed the differences in learning strategies and orientations and conceptions of learning among Dutch and Asian students; and Brito Neto, Smith and Pedersen (2014) analyzed a multicultural group of flight attendants’ perception of e-learning courses, with a particular focus on course relevance and learner motivation, cultural sensitivity, course organization, and course interactivity.

After an extensive review of the literature, Hofstede’s model remains the most widely adopted in the field of education (Viberg & Grönlund, 2013; Nistor, Göğüş & Lerche, 2013). Despite the existence of the updated 6D model, educational researchers still use the outdated 4D/5D model (Marambe et al., 2012; Viberg and Grönlund, 2013; Nistor et al., 2013). This is the second educational contribution of this study.

Finally, from a methodological point of view, multiple correspondence analysis was combined with cluster analysis to categorize international students’ perceptions and to determine the most important variables in online learning that deliver culturally sensitive instruction. Based on the multiple correspondence analysis method, we aim to obtain a global picture of the salient relationships among learner, institutional, and outcome variables.

The cultural dimensions of learning based on the Hofstede’s model: The cases in Spain, the US, Mexico and China

From 1968 to 1972, Professor Geert Hofstede, leaning on his IBM experience, developed a model to characterize national cultures based on four dimensions (Hofstede, 1983). In 1991, based on Michael Harris Bond’s study (Hofstede & Bond, 1988), Hofstede added a fifth dimension, and finally Hofstede transformed his 5D model to a 6D one, which includes the following dimensions (adapted and described for the educational community):

- **Power distance index (PDI).** This dimension refers to the extent to which the members of organizations accept an equal or unequal power distribution. Countries with a high PDI are more likely to accept a hierarchical structure. Meanwhile, in countries with a low PDI, authority is decentralized and leadership is a bit more democratic (Hofstede, 2011).

- **Individualism vs. collectivism (IDV).** Societies with high IDV are considered “behavioral” (McFeeters, 2003). Students expect to be treated as fundamentally equal to peers and faculty and teachers are in charge of stimulating passive learners, using prizes and negative reinforcement or punishment to achieve this goal. Societies with low IDV are considered “constructivist” (McFeeters, 2003). Teachers are merely facilitators in the teaching-learning system, and students show a greater dependence on social relationships.

- **Masculinity vs. femininity (MAS).** This dimension gives details about how an education system can focus on cooperation and security or recognition and advancement (Cambridge, 2012). Countries with a high MAS index encourage competition between students and teachers only reward the excellence. On the other hand, countries with a low MAS index promote a friendly and collaborative learning environment.

- **Uncertainty avoidance index (UAI).** This dimension refers to the degree to which a person feels uncomfortable due to a sense of uncertainty and ambiguity, and it is associated with students’ behaviors toward knowledge construction in the learning context. In cultures with a high level of uncertainty avoidance, students focus on getting the right answer from their teachers. In countries with a low uncertainty avoidance index, there is an open-minded learning process, and learners are allowed to contribute to discussions (Hofstede, 2011).

- **Pragmatic vs. normative (PRA).** A pragmatic pedagogy is based on the “learning by doing” approach. Students establish knowledge by practising by themselves instead of garnering knowledge through repetition. Teachers’ functions are providing learning content and guiding learners to use the research instruments an efficient and effective way. Additionally, a normative approach (a low index score) is based on stable knowledge and rules (Hofstede, 2011); this approach characteristically privileges talents and
aptitudes related to abstract science and theory. Students want immediate gratification and are rarely creative. They attribute success and failure to luck.

- Indulgence vs. restraint (IND). In this dimension, an indulgence score generally implies that the culture allows freedom of speech. A relaxed structure governs the relationship between teachers and learners. The opposite pole of this spectrum relates to restraint. A restraint score implies that learners occupy a subordinate position in relation to teachers, and student motivation tends to be weak in this very stable learning structure.

As will be discussed below, this paper relates the Hofstede cultural dimensions to the cases of the Open University of Catalonia (UOC) in Spain, the University of New Mexico (UNM) in the US, the University of Peking (PKU) in China, and the Autonomous Popular University of the State of Puebla (UPAEP) in Mexico from an educational perspective. These universities were selected for two different reasons. First, all institutions have implemented a learning process of the highest quality for more than 20 years and have more than 10 years of e-learning experience. Second, the studied universities are located in countries that differ significantly in cultural terms.

Although Hofstede’s model has been used as a framework in this study and in numerous cultural research studies of the state of the art, the literature contains several critiques of the model. Most of these studies have identified the conceptualization of the culture as nation as a limitation in the work of Hofstede and the method of data collection or set culture as static rather than as dynamic are some criticisms that Hofstede’s work have received over the years (McLeay & Wesson, 2014). However, in spite of these criticisms, Hofstede’s framework is the most widely accepted in cross-cultural educational studies (Marambe et al., 2012; Viberg & Grönlund, 2013; Nistor et al., 2013). Furthermore, Trompenaars and Hampden-Turner (1998) already detected that the country has the lowest entropy of classification and therefore, country is the major contributor in explaining the cultural orientation of their dimensions. We will also analyze the contribution of the country, but in this case in the creation of clusters to determine the homogeneity of students’ perceptions within the same country.

Materials and methods

Sample description

This research work considered a sample of students from UOC, UNM, PKU, and UPAEP. Data were collected using two surveys. The first survey was sent to participants at the beginning of the course, and the second survey was sent at the end of the course. The surveys and accompanying consent forms were originally written in English then translated into the official language(s) of the universities. They were then developed using Opinio and hosted on the University of New Mexico Health Sciences’ secure application server. Participants were invited via email to respond to the online questionnaire and were given four weeks to respond. One reminder was sent to all respondents after the first two weeks. All factors in the questionnaire were scored on a four-point Likert scale. The reliability results of each of the educational variables are detailed in Appendix 1.

This research study adopted the input-process-output model of learning proposed by Barbera and Linder-VanBerschot (2011). The first survey included five learner (i.e., input) factors:

- General self-efficacy (GSE)
- Self-efficacy online (SEO)
- Motivation (M)
- Prior knowledge (PK)
- The students’ course expectations (CE)

The second survey included eight institutional (i.e., process) factors and three outcome (i.e., output) factors. The eight institutional factors were as follows:

- Learner support (LS)
- Social presence (SP)
- Instruction (I)
- The quality of the learning platform (LP)
- Instructor interaction (II)
- Learner interaction (LI)
Learning content (LC)
Course design (CD)

There were three outcome factors:
- Learner satisfaction (LST)
- Knowledge acquisition (KA)
- Ability to transfer (AT)

Analysis methodology

Data cleaning

The first part of the experiment includes the preprocessing and data cleaning stage. A sample of 1,175 and 709 student evaluations from UOC, PKU, UNM, and UPAEP are analyzed in this study. The first survey has fourteen questions, whereas the second survey has thirty-three questions. Ten of the fourteen questions on the first survey are control questions and are not considered in the study. They were included to determine the validity of the students’ responses. The same holds true for the second survey: only eleven of the thirty-three questions were considered in the study.

Table 1. Number of participants for the first and second surveys

<table>
<thead>
<tr>
<th>University</th>
<th>First survey</th>
<th></th>
<th></th>
<th></th>
<th>Second survey</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ISS</td>
<td>SME</td>
<td>MDR</td>
<td>FSS</td>
<td>ISS</td>
<td>SME</td>
<td>MDR</td>
<td>FSS</td>
</tr>
<tr>
<td>UOC</td>
<td>734</td>
<td>44</td>
<td>23</td>
<td>667</td>
<td>383</td>
<td>118</td>
<td>22</td>
<td>243</td>
</tr>
<tr>
<td>UNM</td>
<td>62</td>
<td>2</td>
<td>7</td>
<td>53</td>
<td>41</td>
<td>15</td>
<td>0</td>
<td>26</td>
</tr>
<tr>
<td>PKU</td>
<td>180</td>
<td>29</td>
<td>24</td>
<td>127</td>
<td>85</td>
<td>16</td>
<td>11</td>
<td>58</td>
</tr>
<tr>
<td>UPAEP</td>
<td>199</td>
<td>0</td>
<td>0</td>
<td>199</td>
<td>200</td>
<td>15</td>
<td>7</td>
<td>178</td>
</tr>
<tr>
<td>Total</td>
<td>1,175</td>
<td>75</td>
<td>54</td>
<td>1,046</td>
<td>709</td>
<td>164</td>
<td>40</td>
<td>505</td>
</tr>
</tbody>
</table>

Note. ISS = initial sample size; MDR = missing data records; SME = single-minded evaluators; FSS = final sample size.

Analyzing the data robustness and quality of the entire dataset is required before performing computational experiments. In this study, 75 of 1,175 students gave the same scores for all fourteen questions on the first survey (including the control variables and the studied variables), and 164 of 709 students provided the same evaluation criteria for all thirty-three questions on the second survey. These evaluators were called single-minded evaluators, and they were removed from the final dataset (as suggested in Fokoue & Gündüz, 2013).

Furthermore, participants with missing values for the study variables were also excluded from the dataset. This exclusion process was considered an adequate strategy because the missing values were assumed to be missing completely at random (Little & Rubin, 2014) and the sample size was still above 250 after the exclusion of participants with missing data in the two surveys. Table 1 describes initial sample size (ISS), missing data records (MDR), single-minded evaluators (SME), and final sample size (FSS).

Multiple correspondence analysis

Multiple correspondence analysis (MCA) is used to uncover a low-dimensional space of multivariate categorical variables where the interdependencies among these categorical data are graphically visualized. MCA is an extension of correspondence analysis, which allows one to analyze the existing relationships between several categorical variables (Le Roux & Rouanet, 2010). In this study, MCA is used to characterize the different elements involved in the learning experience (i.e., learner, institutional, and outcome factors) and to obtain individual scores for the clustering space.

Cluster analysis

Clustering is used to identify relatively homogeneous clusters of student respondents’ profiles. In this work, cluster analysis uses the students’ scores on the first two dimensions of the MCA as input variables. Two clustering solutions are provided (one per survey). In that way, cluster analysis is used to classify students into groups based on the students’ perceptions when they enroll in the course and when they finish it.
Results and discussion

Characterizing students’ perceptions in the studied countries

Analyzing the learner factors

From the multiple correspondence analysis, a two-dimensional solution was considered the most adequate. The first and second dimensions presented have eigenvalues of 2.229 and 1.469, inertia values of 0.446 and 0.294, and Cronbach’s alpha values of 0.689 and 0.399, respectively. Discrimination measures (Figure 1) and a joint plot of category points (Figure 2) were obtained. The most discriminant variables for Dimension 1 were M (0.564), SEO (0.498) and GSE (0.455); the most discriminant variables for Dimension 2 were M (0.397), GSE (0.383) and SEO (0.359) (Figure 1). As seen in Figure 1, the two dimensions are described through a combination of the same three categorical variables (M, GSE, and SEO); PK and CE are the variables with the least discriminant power in the two projections considered.

From the results and their graphical visualization, the two dimensions have been termed “learners’ autonomy,” and the two variables involved are efficacy and motivation, which are actually two of the most important defining variables for the theoretical level of autonomy. Basically, participants are ranked in these two dimensions based on their motivation and self-confidence levels. Students from UOC and UNM are the students with the highest levels of autonomy, followed by students from PKU and UPAEP. Furthermore, in these two dimensions, the variables with the highest correlations are M, SEO and GSE. Bénabou and Tirole (2002) argued that motivation is correlated to self-efficacy, so our findings are consistent with the literature.

China and Mexico, which each have a high power distance index, are more likely to accept a hierarchical structure and their students’ attitudes tend to be more passive, which directly implies lower motivation levels. Meanwhile, in countries with a low power distance index (e.g., Spain or the US), teachers expect students to take initiative; these students are thus expected to be more highly motivated. These expectations are substantiated in Figure 2, where UNM and UOC students are associated with higher motivation values than are PKU and UPAEP students.

![Figure 1. Multiple correspondence analysis learners' dimensions: discrimination measures](image)

UOC and UNM students reported being more motivated than PKU and UPAEP students, which could also be connected to the Hofstede individualism dimension. Kwan and Wong (2015) noted that motivation and achievement are strongly linked to individualism and collectivism, which is also consistent with our results. In our study, individualistic societies (UNM and UOC) tend to have higher motivation values, whereas collectivistic societies (PKU and UPAEP) tend to have lower motivation values. Finally, it is also worth mentioning that most results suggest that there are conditions under which individualists will be more motivated to perform than collectivists and conditions under which collectivists will perform better than individualists.

The universities are ranked with respect to their overall PK as follows: PKU, UOC, UNM, and UPAEP (Figure 2). The same ranking also applies when considering the Hofstede pragmatism dimension. PK and AT are highly
correlated because the more PK the student has on one topic, the easier it is for that student to relate the new knowledge to their own experiences and learn more effectively (Rias & Zaman, 2013). It seems likely that these types of students (PKU students) are prone to have better PK. Conversely, students with normative educations (UPAEP students) are rarely creative and need immediate gratification. In this context, the students with previous normative education should have lower PK than the students with previous pragmatic education, as confirmed in Figure 2.

Finally, individuals with high indulgence scores generally exhibit a willingness to realize their impulses/desires in terms of enjoying life and having fun. They place a higher degree of importance on leisure time and a lower degree of importance to commitment. Hofstede classified Mexico as an indulgent society, whereas the remaining countries qualified as restrained societies (with the US hovering between indulgence and restraint). In general, students in societies with high indulgence scores will tend to have less motivation than students in restrained societies to enroll in and start an online educational course (Figure 2). These findings are aligned with the psychology literature and specifically with the study of Van den Berg (2011) in which commitment is presented as one of several energizing forces for motivated behaviour. This is consistent not only with our teaching experience but also with the statistical data reported by the Mexican students (Figure 2).

Analyzing the institutional factors

As in the previous case, the results from the multiple correspondence analysis are reflected in a two-dimensional figure (Figures 3 and 4). Inertia data (0.506 for Dimension 1 and 0.340 for Dimension 2) and Cronbach’s alpha values (0.861 for Dimension 1 and 0.723 for Dimension 2) can be extracted from this multiple correspondence analysis. The discrimination measurement is illustrated in Figure 3. The most discriminant variables for Dimension 1 were CD (0.649), LC (0.608) and SP/I (0.591), whereas the most discriminant variables for Dimension 2 were CD (0.529), II (0.455), and I (0.490). Thus, LS, LP, and LI are the variables with the least discriminant power in the two-dimensional MCA analysis.

Dimension 1 is defined by institutional variables (CD, LC, and SP), whereas Dimension 2 is defined by instruction variables (CD, I and II). Dimension 1 shows more variance than Dimension 2, which perhaps suggests that students enrolled in online programs place more importance on institution-related factors than on instructor-related factors. In this way, online learning and face-to-face learning seem to have opposing preferences. It makes sense that online students show preferences for institutional variables. A university’s prestige and ranking are among the top variables that students analyze when they are choosing which university to enroll in for the next academic year; online students will most likely never meet their professors, which justifies their preference for institutional factors. The design and structure of the theoretical content (CD and LC) must be thoroughly prepared because e-learning students organize their studies according to their personal circumstances. Furthermore, these findings show that learners want to feel that others appreciate their work (SP). Teachers should not neglect this value. These findings may help teachers to know which (institutional or instructional) aspects have the best value, which depends on the type of education in which they are involved.
The correlations among the categorical variables represented in Figure 3 are analyzed as follows. The first group includes CD and I, and the second group includes LC and SP:

- On the one hand, a positive correlation between CD and I was detected. Most decisions that influence a course’s success take place before a course begins. Therefore, cautious planning during the design stage makes teaching and learning activities easier (Moore, 2013).

- On the other hand, a positive correlation between LC and SP was found. If students are very interested in the learning content, it is very likely that they will feel the need to interact with the learning community to extract information of interest (Shackelford & Maxwell, 2012).

Hereafter, our focal point is Figure 4. As described in the cultural dimension of the learning section, PKU and UPAEP are universities with behavioral educational systems that have lower rates of individualism (see China’s and Mexico’s indulgence Hofstede scores). Recently, online university practitioners have tried to convince society that the constructivist educational model is characteristic of this type of university (Tam, 2000). However, how is it possible that online students in China and Mexico (with behavioral education systems) are satisfied with the constructivist education received? Chinese and Mexican online universities still tend to have a bit of a behavioral flavor (despite great efforts by online university practitioners to impose a constructivist educational model). These behavioral remnants are also understandable, given the nature of online universities and online university students. Despite these inconsistencies, online university students are generally pleased with the education that they receive, given the universities’ proximity to positive assessments.
Analyzing the outcome factors

The outcome factors are also subjected to a multiple correspondence analysis, leading to a two-dimensional graph that summarizes the students’ perceptions. Inertia data (0.647 for Dimension 1 and 0.590 for Dimension 2) and Cronbach’s alpha (0.727 for Dimension 1 and 0.653 for Dimension 2) values can be extracted from this inquiry. Moreover, Figures 5 and 6 show the discrimination measures plot and the conjoint categorical correlation plot, respectively. The most discriminant variables in Figure 5 were LST (0.653) and AT (0.681) for Dimension 1 and LST (0.615) and AT (0.640) for Dimension 2. Thus, the two-dimensional illustration is drawn by combining two elements (LST, AT), and KA has the magnitude with the least impact. From the second plot, an interpretation of points demonstrates how cultural differences are unnoticed after the course is finalized.

The LST and AT variables (which explain the dimensions described in Figure 6) are also highly correlated (as shown in Figure 5). AT is defined as the extent to which students are able to apply their learning to future situations. Instructional activities promote student skills that are related to knowledge transfer. Therefore, according to the empirical results, the instructional activities were correctly carried out in the online universities of this case study. On the other hand, online students’ motivations vary widely. In our case study, it is assumed that learners need to acquire knowledge for work-related reasons. In this sense, students assess very positively the instructional activities that allow them to apply their acquired knowledge in their working lives. This finding could justify the correlation between AT and LST.

The extent to which members of different cultures vary in their reactions to uncertainty may have a major impact on how perceived learning quality affects students’ satisfaction. Students from countries with high uncertainty avoidance levels are comfortable in a learning process that is structured by teachers; in contrast, in countries with low uncertainty avoidance levels, students are comfortable in an open-minded learning environment. Figure 6 shows that students from cultures with high degrees of uncertainty avoidance are slightly less satisfied (students from UPAEP and UOC) than are students from cultures with low degrees of uncertainty avoidance (students from UNM and PKU). This lesser degree of satisfaction occurs when their learning expectations have not been met, perhaps as a result of an educational problem. In light of the tolerance zone concept, this finding suggests a narrower range of acceptable learning outcomes for cultures with high levels of uncertainty avoidance. Therefore, learning quality efforts should be explicitly designed to reflect intercultural differences in the planning and training of service personnel.

These findings also suggest that online learning is more related to open-minded learning than a traditional learning scenario. In this way, American (UNM) and Chinese (PKU) students were acclimatized to flexible structures (according to Hofstede scores), and Spanish (UOC) and Mexican (UPAEP) students were unacclimatized to open-minded training. We observed during our experience in the online programs under study that students with a high level of uncertainty avoidance (Spanish and Mexicans students) needed stricter teaching guidance throughout the course. Moreover, these students were insecure and felt anxious when they needed to make decisions (Hofstede, 2011). We experienced this low level of acclimatization in open-minded environments when group activities were proposed. Students from countries with high UAI (Spanish and Mexicans students)
showed high reluctance toward these activities. Despite these examples, the American and Chinese students appreciated this open-minded approach, as did UOC and UPAEP learners, who acknowledged that enrolling in the course allowed them to transfer their knowledge to their daily routines.

Figure 6. Multiple correspondence analysis outcome dimensions: Joint category plot

Cluster analysis of the student’s perceptions

First survey

Table 2 shows cross-tabulations that relate the four groups with the most discriminant variables (general self-efficacy, self-efficacy online, and motivation) and the four universities (UOC, UNM, PKU, and UPAEP) of the case study. Table 2 shows the distribution of students’ perceptions according to the most discriminant variables and the distribution of students’ perceptions with respect to the universities.

<table>
<thead>
<tr>
<th></th>
<th>C1 (n(%))</th>
<th>C2 (n(%))</th>
<th>C3 (n(%))</th>
<th>C4 (n(%))</th>
</tr>
</thead>
<tbody>
<tr>
<td>General self-efficacy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>6 (14%)</td>
<td>24 (55.8%)</td>
<td>4 (9.3%)</td>
<td>9 (20.9%)</td>
</tr>
<tr>
<td>D</td>
<td>9 (14.3%)</td>
<td>28 (44.4%)</td>
<td>1 (1.6%)</td>
<td>25 (39.7%)</td>
</tr>
<tr>
<td>A</td>
<td>486 (85.1%)</td>
<td>16 (2.8%)</td>
<td>55 (9.6%)</td>
<td>14 (2.5%)</td>
</tr>
<tr>
<td>SA</td>
<td>117 (31.7%)</td>
<td>42 (11.4%)</td>
<td>201 (54.5%)</td>
<td>9 (2.4%)</td>
</tr>
<tr>
<td>Self-efficacy online</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>0 (0%)</td>
<td>26 (59.1%)</td>
<td>5 (11.4%)</td>
<td>13 (29.5%)</td>
</tr>
<tr>
<td>D</td>
<td>3 (7.3%)</td>
<td>16 (39%)</td>
<td>0 (0%)</td>
<td>22 (53.7%)</td>
</tr>
<tr>
<td>A</td>
<td>454 (86.3%)</td>
<td>20 (3.8%)</td>
<td>40 (7.6%)</td>
<td>12 (2.3%)</td>
</tr>
<tr>
<td>SA</td>
<td>161 (37%)</td>
<td>48 (11%)</td>
<td>216 (49.7%)</td>
<td>10 (2.3%)</td>
</tr>
<tr>
<td>Motivation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>4 (7%)</td>
<td>36 (63.2%)</td>
<td>5 (8.8%)</td>
<td>12 (21.1%)</td>
</tr>
<tr>
<td>D</td>
<td>2 (3.6%)</td>
<td>19 (33.9%)</td>
<td>0 (0%)</td>
<td>35 (62.5%)</td>
</tr>
<tr>
<td>A</td>
<td>388 (88.6%)</td>
<td>16 (3.7%)</td>
<td>28 (6.4%)</td>
<td>6 (1.4%)</td>
</tr>
<tr>
<td>SA</td>
<td>224 (45.3%)</td>
<td>39 (7.9%)</td>
<td>228 (46.1%)</td>
<td>4 (0.8%)</td>
</tr>
<tr>
<td>Universities</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UOC</td>
<td>493 (73.9%)</td>
<td>9 (1.3%)</td>
<td>165 (24.7%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>UNM</td>
<td>29 (54.7%)</td>
<td>1 (1.9%)</td>
<td>23 (43.4%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>PKU</td>
<td>73 (57.5%)</td>
<td>5 (3.9%)</td>
<td>35 (27.6%)</td>
<td>14 (11%)</td>
</tr>
<tr>
<td>UPAEP</td>
<td>23 (11.6%)</td>
<td>95 (47.7%)</td>
<td>38 (19.1%)</td>
<td>43 (21.6%)</td>
</tr>
</tbody>
</table>

Note. SD = strongly disagree; D = disagree; A = agree; SA = strongly agree.

The elbow criterion determined the optimal numbers of clusters, i.e., four for the first survey (C₁, C₂, C₃, C₄). Each group corresponds to an autonomy level: a very high level of autonomy, a high level of autonomy, a medium level of autonomy, and a low level of autonomy. C₂ includes students with the lowest autonomy level; C₄ includes students with the second lowest autonomy level; C₁ includes students with a high autonomy level; and C₃ includes the students with the highest autonomy level. Furthermore, cross-tabulations also allow us to...
associate a level of the most discriminant variables with an autonomy level. For example, the 55.8% of students with the lowest general self-efficiency values are included in the second cluster, which groups students with the lowest possible autonomy values.

Universities are also ranked according to the degree of their students’ autonomy. Thus, the US ranks first (UNM is the university with the greatest number of students with a very high autonomy levels and the second greatest number of students with high autonomy levels); Spain ranks second (73.9% of UOC students have high levels of autonomy); and China and Mexico rank third and fourth, respectively.

**Second survey**

Table 3 is a cross-tabulation table that includes the following factors: the three groups (C₁, C₂, C₃) that are determined by the elbow criterion, the most discriminant variables for the second survey (instruction, learner content, course design, learner satisfaction, and ability to transfer), and the four online universities. In this case, universities and students are ranked according to their levels of satisfaction.

The elbow criterion procedure determined that the optimal number of groups for the second survey was three. The three groups relate to the students’ different levels of satisfaction: very satisfied, satisfied, and unsatisfied. The groups rank as follows: C₁ includes the unsatisfied students; C₂ includes the satisfied students; and C₃ includes the unsatisfied ones. Furthermore, Table 3 correlates the grade of each variable (SD, D, A, SA) with the group, which allows us to know the proportion of students that agrees or disagrees on a specific variable included in each group. For example, the 67.6% of students who are highly satisfied with instruction (SA/I) are included in C₃. Table 3 also allows us to determine the most important variables that characterize the final satisfaction level. For example, learners from asynchronous online systems evaluate the course design (83.4%) and learning content (80.7%) more positively than the instruction (67.6%).

**Table 3. Cluster analysis second survey**

<table>
<thead>
<tr>
<th></th>
<th>C₁ (n(%)</th>
<th>C₂ (n(%)</th>
<th>C₃ (n(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instruction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>18 (90%)</td>
<td>2 (10%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>D</td>
<td>18 (20.5%)</td>
<td>68 (77.3%)</td>
<td>2 (2.3%)</td>
</tr>
<tr>
<td>A</td>
<td>6 (2.3%)</td>
<td>208 (80.6%)</td>
<td>44 (17.1%)</td>
</tr>
<tr>
<td>SA</td>
<td>1 (0.7%)</td>
<td>44 (31.7%)</td>
<td>94 (67.6%)</td>
</tr>
<tr>
<td>Learner content</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>7 (70%)</td>
<td>3 (30%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>D</td>
<td>22 (53.7%)</td>
<td>19 (46.3%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>A</td>
<td>14 (4.5%)</td>
<td>273 (86.9%)</td>
<td>27 (8.6%)</td>
</tr>
<tr>
<td>SA</td>
<td>0 (0%)</td>
<td>27 (19.3%)</td>
<td>113 (80.7%)</td>
</tr>
<tr>
<td>Course design</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>9 (100%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>D</td>
<td>21 (55.3%)</td>
<td>17 (44.7%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>A</td>
<td>12 (3.9%)</td>
<td>281 (91.5%)</td>
<td>14 (4.6%)</td>
</tr>
<tr>
<td>SA</td>
<td>1 (0.7%)</td>
<td>24 (15.9%)</td>
<td>126 (83.4%)</td>
</tr>
<tr>
<td>Learner satisfaction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>3 (100%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>D</td>
<td>18 (50%)</td>
<td>17 (47.2%)</td>
<td>1 (2.8%)</td>
</tr>
<tr>
<td>A</td>
<td>17 (7%)</td>
<td>214 (87.7%)</td>
<td>13 (5.3%)</td>
</tr>
<tr>
<td>SA</td>
<td>5 (2.3%)</td>
<td>91 (41%)</td>
<td>126 (56.8%)</td>
</tr>
<tr>
<td>Ability to transfer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>9 (75%)</td>
<td>3 (25%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>D</td>
<td>18 (26.9%)</td>
<td>48 (71.6%)</td>
<td>1 (1.5%)</td>
</tr>
<tr>
<td>A</td>
<td>12 (3.9%)</td>
<td>242 (79.1%)</td>
<td>52 (17%)</td>
</tr>
<tr>
<td>SA</td>
<td>4 (3.3%)</td>
<td>29 (24.2%)</td>
<td>87 (72.5%)</td>
</tr>
<tr>
<td>Universities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UOC</td>
<td>22 (9.1%)</td>
<td>156 (64.2%)</td>
<td>65 (26.7%)</td>
</tr>
<tr>
<td>UNM</td>
<td>2 (7.7%)</td>
<td>8 (30.8%)</td>
<td>16 (61.5%)</td>
</tr>
<tr>
<td>PKU</td>
<td>1 (1.7%)</td>
<td>41 (70.7%)</td>
<td>16 (27.6%)</td>
</tr>
<tr>
<td>UPAEP</td>
<td>18 (10.1%)</td>
<td>117 (65.7%)</td>
<td>43 (24.2%)</td>
</tr>
</tbody>
</table>

**Note.** SD = strongly disagree; D = disagree; A = agree; SA = strongly agree.

Finally, universities can also be ranked according to their students’ levels of satisfaction. Because 98.3% of its students are either satisfied or very satisfied with the courses and their relevance, PKU is the university with the most satisfied students. With a low percentage of totally unsatisfied students (only 7.7% of its students are unsatisfied), UNM ranks second. In Spain, 9.9% of UOC students are absolutely unsatisfied, but 90.9% of
students are satisfied or very satisfied with their course. Lastly, 10.1% of UPAEP students in Mexico are very dissatisfied with their course.

**Analyzing the limitations of the study**

As explained earlier, the main limitation of Hofstede’s cultural model is the conceptualization of the culture as nation, i.e., the culture within the same country is assumed to be homogeneous. An advantage that the methodology adopted offers us is the possibility of checking this assumption. If the majority of students in the same country are grouped within the same educational clustering (note that those clusters were created considering only the students’ perceptions and without considering their nationality), then we can affirm that there is a high degree of homogeneity in the students’ perception of that country.

Tables 2 and 3 report students’ nationalities and show how students are grouped according to their learning perceptions. As shown in Table 2, Mexican students reported the highest degree of heterogeneity in their perceptions before enrolling in the course (the majority cluster grouped the lowest percentage of students, 47.7%, when compared with the majority clusters of the remaining countries, UOC 73.9%, UNM 54.7% and PKU 57.5%). On the other hand, during and after the course (Table 3), the most heterogeneous students were the Americans (with the 61.5% of students included in C3).

It is important to mention that if nationality were not an influential variable for the creation of groups, the percentage of students in the majority group in Table 2 would be 25% (four classes were considered) while for the second case (Table 3) would be 33.3% (three classes were considered). As can be seen in Tables 2 and 5, the majority groups in the most heterogeneous countries in both cases are almost double these percentages. Furthermore, the communities of the universities under study are generally composed of students and professors born in different cities (especially the UOC and UNM universities). Taking these two facts into account, we conclude that the conceptualization of culture as nation for the case of the students’ perceptions towards online learning is not a weak assumption for the countries considered.

**Conclusions and future work**

This paper analyses students’ perceptions in four different online learning universities. As a result, the most important institutional and learner factors (both intrinsic and extrinsic values) in e-learning environments are identified. Furthermore, Hofstede’s cultural framework was used to justify the cultural differences among the countries under study. Thus, (1) the dimensions implicated in the learners’ factors were the Power Distance, Individualism, Pragmatic, and Indulgence dimensions; (2) the Individualism dimension was associated with the institutional factors; and (3) the Uncertainty avoidance dimension was connected with the outcome factors.

On the other hand, the multiple correspondence analysis has allowed us to graphically visualize the cultural differences among the different countries during the course. At the beginning of the educational process, the cultural differences were very noticeable. During the course, the four countries were clustered into two groups (a group composed of Spanish, Chinese, and Mexican students and another group composed of American students). Finally, at the end of the course, the students’ perceptions in all countries were very uniform.

Thus, instructional designers and practitioners could benefit from the results of this research.

- For designers: It is important to know how to design courses that truly help people learn. Identifying the critical learning factors for each culture should guide the implementation of online learning environments that include and respect the particular perspectives of international students.
- For teachers: To deliver culturally sensitive instruction, it is essential to be aware of differences among international students’ perceptions. Flexibility in teaching activities, including a wide range of learning activities, is crucial.

During the study, we determined the most influential variables of the whole online educational procedure. This could help practitioners and researchers to detect those educational areas of interest that deserve special attention in online environments that have students from different cultural backgrounds. Hence, we will analyze those variables in depth in future work. From a different perspective, the inclusion of instructors’ perceptions in the analysis and their differences regarding learners’ perceptions will be also considered in future research.
Acknowledgements

The research work of P. Gómez-Rey has been supported with a doctoral grant from the Universitat Oberta Catalunya. Likewise, the research work of F. Fernández-Navarro was partially supported by the TIN2014-54583-C2-1-R project of the Spanish Ministry of Economy and Competitiveness (MINECO), FEDER funds and the P2011-TIC-7508 project of the “Junta de Andalucía” (Spain).

References


Appendix 1

Table 4 includes the items defining the educational variables mentioned in the study. Content validity and reliability were checked at the beginning of the surveys to ensure the readability and clarity of the questionnaire items. The values of Cronbach’s alpha for the first and second survey were 0.84 and 0.86, respectively, which suggest that the questionnaires had adequate reliability.

Table 4. Instrument used in the study

<table>
<thead>
<tr>
<th>First survey</th>
<th>Second survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE</td>
<td>LS</td>
</tr>
<tr>
<td>I am confident that I can effectively deal with any unexpected event (personal or academic) during the course.</td>
<td>I received the technical support I needed when I had a problem.</td>
</tr>
<tr>
<td>SEO</td>
<td>SP</td>
</tr>
<tr>
<td>I am capable of learning in online educational environments.</td>
<td>I felt I was a part of a community of learners in this course.</td>
</tr>
<tr>
<td>PK</td>
<td>I</td>
</tr>
<tr>
<td>I feel motivated to learn in this course.</td>
<td>The instructor used effective teaching strategies.</td>
</tr>
<tr>
<td>M</td>
<td>LP</td>
</tr>
<tr>
<td>I am able to apply information I have learned in other courses to this course.</td>
<td>All-important site content was easy to locate and identify.</td>
</tr>
<tr>
<td>CE</td>
<td>II</td>
</tr>
<tr>
<td>The course information I received before enrolling gave me an accurate picture of the course.</td>
<td>The instructor provided individualized guidance that met my needs.</td>
</tr>
<tr>
<td></td>
<td>LI</td>
</tr>
<tr>
<td></td>
<td>LC</td>
</tr>
<tr>
<td></td>
<td>CD</td>
</tr>
<tr>
<td></td>
<td>LST</td>
</tr>
<tr>
<td></td>
<td>KA</td>
</tr>
<tr>
<td></td>
<td>AT</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Integration of Mobile AR Technology in Performance Assessment

Chao Kuo-Hung, Chang Kuo-En, Lan Chung-Hsien, Kinshuk and Sung Yao-Ting

1National Taiwan Normal University, Taiwan // 2National Taiwan Normal University, Taiwan // 3Taoayun Innovation Institute of Technology, Taiwan // 4Athabasca University, Canada // 5National Taiwan Normal University, Taiwan //

*Corresponding author

(Submitted January 22, 2015; Revised October 26, 2015; Accepted March 8, 2016)

ABSTRACT

This study was aimed at exploring how to use augmented reality (AR) technology to enhance the effect of performance assessment (PA). A mobile AR performance assessment system (MARPAS) was developed by integrating AR technology to reduce the limitations in observation and assessment during PA. This system includes three modules: Authentication, AR Context Aware, and AR Interaction. It identifies the character of system users, recognises subjects and activities that users should join in, and follow and assists users in showing their works and doing assessment anywhere anytime. Both qualitative and quantitative research were applied in this study, and it presents an action research conducted in PA institutions that, through a questionnaire, gave information in connection to the valuation of mobile AR technology and the achieved effects. The proposed mechanism of PA emphasises the approaches adopted to present student works and provides opportunities for enhancing student communication and interaction. In addition, the system enables students to explain their works and incorporate the feedback they receive into future work. More importantly, mobile AR can be applied to offer personalised features and appropriate information in particular areas. From the observation of survey results, students can interact with real or virtual information based on their needs. During this process, students can observe their own works from varying perspectives, acquire vital knowledge, develop the skills of critical thinking, and transform the process into a substantial self-established learning process.

Keywords

Augmented reality, Performance assessment, Mobile learning

Introduction

Based on the popular educational philosophy of allowing students to develop diverse capabilities and achieve active knowledge building, performance assessment (PA) should be considered a vital link in teaching. In addition to assigning a final score to students, the purpose of assessment is to develop a highly in-depth understanding regarding the process that students undergo during learning and provide feedback to assist in student growth. Toptas (2011) indicated that an effective evaluation of the students who answered the questions in a particular period of time will be insufficient. If we want to correct this weakness, the performances of the students must be measured with the observation of the process as well. O’Neill and Osif (1993), and VanTassel-Baska (2014) indicated that assessment plays a vital role in teaching and that the process of assessment consists of goal setting, data collection, organisation, and result analysis. The results can be used to enhance teaching and report the actual progress of students. Turgut and Baykul (2012) point out that the process can be measured alongside the results of learning outputs by measuring the performances. In addition, it is asserted that the measurement of students’ performance gives them the opportunity to effectively learn the concepts, complex events, and their structures.

Nevertheless, a major problem encountered by the education community is determining the appropriateness of educational evaluation. PA has been recognised as one of the most effective methods for assessing this type of high-level thinking because this approach emphasises the application and demonstration of abilities in problem-solving situations and the complexity of problem-solving processes (Wiggins, 1993; VanTassel-Baska, 2014). Previous studies (Bay, Küçükoğlu, Kaya, Gündoğdu, Köse, Ozan, & Taşgun 2010; Jiang, Smith, & Nichols, 1997) have indicated that the primary limitations and disadvantages of the PA approach include the lack of comparison, limited reliability, unsatisfactory economic performance, and low validity. However, the majority of these factors can be attributed to the subjective consciousness of the assessors and errors in the measured situations. By contrast, augmented reality (AR) technology can be employed to display, in real situations, real-time information that is necessary for assessing or learning. From the perspective of cognitive psychology, this approach can be applied to reduce the errors resulting from the process of PA and to minimise the time and economic costs that teachers must bear when observing student behaviour. Therefore, we examined the meaning, relevant studies, and limitations of PA before investigating the effects that incorporating AR technology exert on...
improving PA systems. Subsequently, we applied an AR-based PA system to a cooking course to explore the effects of the application. The results of this study can serve as a reference for implementing PA in teaching.

Performance Assessment (PA)

Performance Assessment (PA) requires students to apply the knowledge and skills they have learned to perform hands-on practice rather than simply revalidating and recollecting the experience of learning (VanTassel-Baska, 2014). This assessment method satisfies the needs of the current trend of constructivist learning and teaching (Chang, 2002). PA motivates students to integrate the knowledge, skills, and dispositions required in the subject, and the results of the assessment can reflect students’ problem-solving abilities in real life and the interest and needs of the students. Performance assessments, which can be conducted to evaluate high-level cognitive abilities and the dispositions and skills of students, are more comprehensive compared with conventional paper-based tests. When evaluating a student for periodic checks or a promotion, there has to be a list of measurable performance criteria that can be applied consistently to all members of a particular class. (Kirovska & Qoku, 2014). Therefore, by adopting specific performance and assessment criteria, teachers can provide students with specific feedback, which motivates students to take initiatives in learning and assume the responsibility to critique their own works and strategies. Therefore, PA is an effective approach for facilitating teaching and learning.

Students who acquire a comprehensive set of motor skill proficiencies and continually engage their skills are more likely to become fit adults (Kalaja, Jaakkola, Luikkonen, & Watt, 2010). Strong factors include competence at performing and confidence in using motor skills, both of which are established through early experiences in physical activity and sport (Rink & Hall, 2008).

The purpose of PA is using assessment to promote student development. During assessment, allowing other students and teachers to provide feedback on the assessment standards, record the processes, and evaluate the progress made by the student under assessment offers more opportunities for teacher–student and peer interactions. During these interactions, students can communicate and explain their learning experience and contemplate the learning process. In addition, the process of teamwork enables students to develop the abilities to communicate and cooperate with their peers and to develop favourable work attitudes. The difference between teachers (experts) and novices is that experts understand how to effectively use the knowledge they have acquired. PA enables teachers to comprehend the thinking process students undergo by observing the process of students’ operations. Thus, teachers can understand whether the students’ operations comply with the prescribed procedure.

Limitation of performance assessment in learning

PA, where the rating is often performed by professionals based on their observation and judgment, is subjective, demanding, and low in reliability compared with paper-based tests. Therefore, a critical problem that necessitates solution in implementing PA is devising fair and objective rating criteria that are easy to apply and can be used to provide feedback to students (Lu, Chen, & Wu, 2005; VanTassel-Baska, 2014).

A previous study regarding PA found that errors in the generalisation of PA are primarily affected by the following four factors: (a) the items or activities used in the assessment, (b) the assessors, (c) the situations in which the assessments are conducted, and (d) the unintentional influence of assessees or other people (Jiang, Smith, & Nichols, 1997).

To determine whether a student has mastered a skill, assessors must collect performance data on multiple occasions. They can obtain all details by conducting a minimal number of observations, thereby reducing the cost of assessment. Simultaneously, assessees can benefit from fair assessments performed based on records that contain all details regarding their performance. When PA is applied to assess the operation and production of actual works, the fairness, objectivity, convenience, and timeliness of assessment must be considered to overcome the limitations of this method. Table 1 lists the primary limitations of PA and how AR technology was employed to solve the problems.

The assessment conducted in implementation activities are considered assessments conducted during activities. PA is often based on observation; thus, it can also be referred to as the work evaluation method. However, the traditional assessment methods have been replaced by the evaluation activities based on the authentic
apprehension (Bay et al., 2010). The concept of assessment brought by the constructivist learning approach must be varied according to the traditional methods, handled as an element of the learning-teaching processes (not independent from the process), and be a part of educational activities.

<table>
<thead>
<tr>
<th>Table 1. Limitations of PA and solutions</th>
<th>How AR solve the problems</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subjective judgments of the evaluator:</strong></td>
<td>The process of student performance can be recorded, and students and teachers can employ AR technology during the rating process. Thus, the assessors and assesses can appear in real-time situations and serve as the direct references for assessment processes, thereby enhancing the accuracy of assessment.</td>
</tr>
<tr>
<td>The results of PA are often affected by the subjective judgments of teachers; additionally, the criteria employed are occasionally confusing, thereby increasing the difficulty involved in comparing and interpreting the assessment results.</td>
<td></td>
</tr>
<tr>
<td><strong>Limited reliability:</strong></td>
<td>We can employ AR to present the processes of work production or the implicit details hidden in the works. Thus, assessment accuracies can be increased substantially and the risks of rating errors resulting from assessor negligence or excessively short observation time can be reduced.</td>
</tr>
<tr>
<td>The majority of manual assessment methods are subject to the subjective influence of the assessors. Unlike standardised tests, for which computer scoring can be adopted, PA relies on assessor observation and judgment. Consequently, the reliability of the assessors should not be overlooked. The errors in assessor reliability result from the assessors, and a satisfactory rating system can reduce assessor errors.</td>
<td></td>
</tr>
<tr>
<td><strong>Unsatisfactory economic performance:</strong></td>
<td>Using AR to present the production processes of works can reduce the travel costs that teachers would otherwise spend for conducting on-site observations. In addition, the assessors can watch videos repeatedly to reduce rating errors and the amount of effort that teachers must spend on assessments. Furthermore, occasionally teachers must simultaneously observe multiple students, thereby rendering them unable to observe all details within a particular period. Adopting AR can prevent this problem.</td>
</tr>
<tr>
<td>The costs spent on PA are considerably greater than those spent on paper-based tests.</td>
<td></td>
</tr>
<tr>
<td><strong>Low validity:</strong></td>
<td>An AR-integrated system can show in real-time the rating standards and the feedback from teachers or peers; thus, the associated cognition of feedback materials and student works can be enhanced. Therefore, students can more effectively immerse themselves into the teaching situations, thereby improving the validity of assessments.</td>
</tr>
<tr>
<td>In PA, ambiguous problem situations test the high-level thinking abilities of assesses. Nevertheless, the validity of ambiguous problems is difficult to control; consequently, the assessment can be irrelevant to the teaching contents.</td>
<td></td>
</tr>
</tbody>
</table>


**Mobile augmented reality**

AR enables users to visualise environments in a real world with the digital information overlaid on real environments (objects or locations), thereby improving user experiences (Berryman, 2012). The combination of additional information and real situations can enhance the senses of reality and presence for people. The theoretical basis for the mobile AR system that integrates human–computer-context interactions is situated cognition. The fundamental argument of the theory is that knowledge acquisition and learning occurs after people interact with situations, which include social environments such as people and social culture, and physical environments such as the contexts formed by scenes and artefacts (Greeno, Collins, & Resnick, 1996). In addition, human–computer-context interactions are difficult. Participants may focus excessively on human–computer interactions and overlook human–context (objects in scenes and information contexts) interactions, which are more crucial than human–computer interactions in real situations. Therefore, the link between additional information and real environments should be emphasised in the virtual information presented in AR (Chang, Chang, Hou, Sung, Chao, & Lee, 2014).

In informal learning, the application of mobile devices has recently attracted an increasing amount of attention (Kwak & Stoddard, 2004; Cabrera et al., 2005). However, studies regarding the application of AR navigation are scant (Damala, Marshal, & Houlier, 2007; Portalés & Perales, 2009). The mixed reality spectrum (Figure 1)
developed by Milgram, Takemura, Utsumi, & Kishino (1994) offers a valuable basis for exploring the integration of reality and virtual reality. AR is situated on the spectrum between virtual and real environments.

**Figure 1.** The mixed reality spectrum and mobile AR technique (Milgram, Takemura, Utsumi & Kishino, 1994)

AR enables users to visualise real environments in a real world with digital information overlaid on actual environments (objects or locations), thereby improving user experience (Berryman, 2012). Azuma (1997) defined three criteria for AR: (a) the combination of virtual and real environments, (b) real-time interaction, and (c) 3D referencing. Scholars generally agree that AR can be used to enhance the experience that users have when interacting with real environments. In addition, virtual information enables users to obtain information that otherwise cannot be directly acquired from the real world. Because of this feature, AR is considered an effective tool that users can employ to achieve objectives in the real world (Azuma, 1997). Figure 2 shows how to use AR to display appropriate information.

**Figure 2.** The application of AR technology

As showing in Figure 2-1, for the purpose of locating virtual information in the right place in real word, tags or markers are necessary for recognition. AR recognises the tag and gets its position as the position of the corresponding virtual information. Figure 2-2 shows how outdoor AR uses GPS location information and compass direction as the tag to locate virtual information.

Since the 1990s, AR has been applied in various fields, including geography (Portalés et al., 2010; Priestnall & Cowton, 2009), linguistics (Liu, 2009), social sciences (Mathews, 2010), mathematical sciences, natural sciences, biomedicine, arts and humanities, leisure and recreation, and advertising and marketing.

In summation, to design a PA learning system that achieves human–computer-context interactions, we employed AR technology to develop a PA system that enables peer assessment. We predefined the criteria on which learners produce their works or assessors evaluate the works in this system. Thus, students can evaluate their own works or peers’ works based on sufficient information, thereby developing strong learning motivations and achieving great efficacy. In addition, teachers or assessors can spend comparatively less time and simultaneously evaluate assessees’ works accurately and fairly.

**System realisation and illustrative example**

**System architecture**

In the field of education, numerous situations cannot be experienced or represented in the classroom setting. AR is the most appropriate technology for incorporating or adjusting students’ learning experience based on specific
needs. AR is defined as a real-world environment whose elements are built upon computer-generated sensory input such as sound, video, graphics, or GPS data. In this study, AR allows students to see virtual objects about peers’ works or contents in a real-world environment with the aid of camera during the assessment process. The overall framework of the use of the mobile AR technique in PA is described in Figure 3.

The entire processes of learning and assessment can be divided into the three modules: the authentication module, the context-aware module, and the interactive assessment module. In addition, the process is supported by three databases on cloud servers, which are student profiles, the AR and virtual object database, and the assessment database. The authentication module enables authorised people to obtain appropriate information for completing corresponding tasks. In the AR context-aware module, mobile devices list appropriate learning contents after detecting users’ cooking subjects. Thus, learners can select appropriate learning materials from the databases. The authoring tool retrieves appropriate materials from the virtual object database before providing them to assessors. AR interaction module assists users in showing their works and doing assessment. Then they can discuss in this system anywhere anytime and give feedback.

For teachers and students, the authentication module shortens the gaps between learners and assessees. The AR context-aware module is situated cognition by using AR to integrate human–computer-context interactions, and it reduces the assessment effort while assessors assess every assesseee at the same time. The AR interaction module assists teachers and students in sharing and discussing feedback more easily and frequently.

Subsequently, learners can use the authoring tool provided by the system to establish an AR marker, work descriptions, and an AR context object. Once all steps are completed and the information is uploaded to the system, appropriate virtual information is used through AR technology to overlay images onto corresponding objects in the real world. Thus, assessors can rate the works conveniently and accurately. Through the AR work presentation technology employed in the system, the assessment module enables assessors to conveniently and directly observe the works of peers. Hence, assessors can provide feedback for the peers they evaluate. In addition, the system can be employed to develop a work-specific exhibition situation for peer references, thereby enabling peer assessors to provide feedback. Additionally, teachers can use the AR PA system to understand the peer assessment performed by students before providing feedback for the assessors and assessees. More importantly, teachers can integrate previous cases to develop new teaching situations that are highly appropriate and inspiring.

**Walk-through illustrative example**

The methods for conducting PA are diverse, including observation, document records, and real-time performance. The methods adopted in this study were real-time performance and peer assessment. Peers who possessed similar knowledge levels observed and learned from each other before offering recommendations. Specifically, a class of 50 sophomore students at the culinary department of a technical institute were recruited as the participants of this study. A PA experiment was conducted during a training course for cooking licences in Western cuisine. The students were divided into groups of five, obtaining a total of ten groups. The group members divided the labour between themselves. The students were randomly assigned to the groups without considering sex or cooking skills. During class, the teacher designated an item from the licensing examination as
a task. The teacher demonstrated the cooking procedure once, after which the completed set was recorded and used as an item marker in the AR PA system. The procedure of the experiment is shown in Figure 4.

Figure 4. The procedure of the experiment

Students first choose the character (assessor or assesse) they want to play. Before students begin the performance, they use mobile devices (tablet and personal computers) to photograph the sample. Subsequently, the system displays real-time information (learning mode) that corresponds to the dish onto the dish image, such as the ingredients that should be prepared and the steps of cooking. Thus, students can follow the instructions during the performance or cook by themselves and record the process using mobile devices. Then the videos are uploaded to cloud servers. After all the dishes are completed, the system integrates all the data for peer assessment. The system enables assessors to review the records of assessed dishes for reference. In the assessment mode, the system lists the content and assessment criteria of the set for assessees. Assessees prepare the ingredients for cooking the dishes and record every step. Finally, the system integrates all of the information for the teachers to provide ratings and feedback.

During assessment, the AR PA system identifies each dish and lists the contents when assessors use the cameras on their mobile devices to photograph the sets completed by assessees. When an assessor selects the name of a particular dish, the video showing the cooking process is immediately shown on the screen. In addition, the assessment criteria are displayed simultaneously, enabling the assessors to perform the assessment intuitively and clearly. Because the entire cooking procedures were videotaped, the assessors were able to observe all the details that interested them. Thus, the assessors did not miss crucial details as they otherwise would have when simultaneously observing several groups of students. Additionally, the assessment criteria adopted are consistent because they are shown in real time. Hence, the errors in PA can be minimised.

Furthermore, the assessors can provide real-time feedback and recommendations during assessment. The feedback can be uploaded to cloud servers immediately following assessment. Thus, the students can immediately review and share the feedback and recommendations regarding their works and further discuss among themselves by using the system. Real-time sharing and the real-time display of assessment criteria enable students to immediately understand the advantages and disadvantages of their works and to use the feedback to improve their works. Thus, the learning objectives were achieved.
**Procedure and evaluation**

**Setting and participants**

In this study, we wanted to know whether the effect of PA can be enhanced by AR system. Could AR technology solve the problems of PA? Thus, both qualitative and quantitative research designs were used in this study. To validate the proposed framework and the effectiveness of the system, we conducted a survey. In the study, a class of 50 sophomore students in the culinary department of a technical institute were recruited as the participants of this study in Taiwan. And 47 survey results were recorded from these 50 participants. According to Van Zundert et al. (2010), peer assessment is facilitated by working in small groups of three or four students. These students are better able to compare feedback from different peers to determine its relevance. The grouping size of a cooking team cannot be small, and there must be an odd number of participants for the sake of voting convenience when there is a decision to make. A five-person grouping is the closest grouping size that fits all the situations. Therefore, in our study, the students were arranged in ten groups of five students. The experiment arranged two activities of PA: pen-and-paper assessment and mobile AR assessment. All groups undertook peer assessment on paper first, while the same groups employed the system to do peer assessment one week later. Finally, a survey considering the amount of time spent focusing on peer discussion and students’ attitude regarding the use and acceptance of MARPAS was assigned to students to evaluate the effectiveness of the system. The grades of assessment were compared between assessment by paper-and-pen and assessment using MARPAS.

During the experiment, all participants had to complete the training of peer assessment in order to understand the process of observation, assessment, and interaction. Subsequently, each group joined two activities: assessment on papers and assessment on mobile devices by using MARPAS. The teacher demonstrated the cooking procedure once, after which the completed set was recorded and used as an item marker in the AR PA system. The AR application constructed the relation between the image of the draft and the exposition. However, assessment criteria defined by the teacher were different based on the various dishes. The assessment criteria represented the teacher’s requests and were also the basis for marking peers’ works.

After students filled out the questionnaire, interviews were arranged with each group, where the aim was to explore the participants’ attitudes toward the strengths and limitations of the system and any suggestion students may have for improvement. And the results of assessments were recorded to measure the progress of students’ works. A qualitative approach was used to analyse the participants’ feedback. Then, the interviews about MARPAS were arranged with teachers who assigned these performance assessments.

**Data collection and analysis**

After the experiment, the assessment process and results were recorded. In order to validate the proposed framework, the experimental results were analysed. In addition, a survey was conducted to collect additional data from participants. A structured questionnaire was developed for this purpose and was sent to five experts to evaluate reliability and validity. These experts included two professors in cognitive psychology, one professor in computer science, and two cooking teachers. Subsequently, the questionnaire was distributed to each student in the ten groups. This questionnaire contained three sections: the first section related to the experience in joining peer assessment, while the second section dealt with the feedback in using the mobile service. The third section figured out how AR technology assisted students in demonstrating their work. The reliability of this questionnaire was checked by testing the value of Cronbach’s alpha. Cronbach’s alpha of the questionnaires was estimated as internal consistency reliability. The results showed that the reliability of the questionnaires was good (whole questionnaire: alpha = 0.937, section 1: alpha = 0.925, section 2: alpha = 0.896, section 3: alpha = 0.781). All show that the questionnaire is reliable.

In the questionnaire, there were three questions pertaining to personal information, 16 questions for the first section, 8 questions for the second section, 5 questions for the third section, and 4 questions for the fourth section. The questionnaire used five-point Likert scale to register students’ answers. Collected data was analysed using correlations, associations, and descriptive statistics in order to assess the relationships between variables.
Findings and discussion

According to Kothari (2009), for sample sizes of more than 30, the $t$-distribution is so closed to the normal distribution that one can use to approximate the $t$-distribution (Kothari, 2009). Therefore, the $t$-test was used to validate the effectiveness of the system. Seven items in the first section of the questionnaire were selected to analyse students’ attitudes toward the use of peer assessment. The hypotheses are listed below:

$H_0$: Students did not agree on effect of performance assessment ($\mu \leq 3$)

$H_1$: Students agreed on effect of performance assessment ($\mu > 3$)

Table 2. Analysis results for the effect of PA

<table>
<thead>
<tr>
<th></th>
<th>Test value = 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t$</td>
</tr>
<tr>
<td>I could understand the teacher’s requests in this course more clearly from performance assessment activities.</td>
<td>8.076</td>
</tr>
<tr>
<td>I could understand peers’ recognition of my work from peer assessment activities.</td>
<td>9.913</td>
</tr>
<tr>
<td>Performance assessment activities increased my learning motivation.</td>
<td>11.524</td>
</tr>
<tr>
<td>Performance assessment activities engaged my attention in the course.</td>
<td>8.398</td>
</tr>
<tr>
<td>Performance assessment activities increased the interaction with the teacher.</td>
<td>9.141</td>
</tr>
<tr>
<td>Performance assessment activities increased the interaction with peers.</td>
<td>10.960</td>
</tr>
<tr>
<td>The suggestions from peers were helpful to me.</td>
<td>10.446</td>
</tr>
</tbody>
</table>

The second section of the questionnaire related to students’ attitudes toward the use of mobile service for PA. The hypotheses are listed below:

$H_0$: Students did not agree on effect of mobile service for performance assessment ($\mu \leq 3$)

$H_1$: Students agree on effect of mobile service for performance assessment ($\mu > 3$)

The experimental results reported in Tables 2 show that students agreed on effect of mobile service, which made PA activities more convenient and fairer.

Table 3. Analysis results for the effect of mobile service

<table>
<thead>
<tr>
<th></th>
<th>Test value = 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t$</td>
</tr>
<tr>
<td>Using mobile service could reduce the cost of traffic during performance assessment.</td>
<td>6.948</td>
</tr>
<tr>
<td>Performance assessment activities could progress anytime anywhere by using mobile service.</td>
<td>9.936</td>
</tr>
<tr>
<td>It is fair to assess peers’ works regardless of the relationship with peers by using mobile service.</td>
<td>6.933</td>
</tr>
<tr>
<td>The limitation of the hardware location could be eliminated by using mobile service.</td>
<td>5.657</td>
</tr>
</tbody>
</table>
The third section of the questionnaire clarifies whether the use of AR technology to demonstrate students’ works during mobile PA is helpful. The section includes five items and associated hypotheses are listed below:

\( H_0: \) AR technology did not assist the activities of mobile performance assessment \( (\mu \leq 3) \)

\( H_1: \) AR technology assists the activities of mobile performance assessment \( (\mu > 3) \)

The experimental results in Table 3 reveal that participants thought using AR technology to enhance the work demonstration is helpful for students in the course.

<table>
<thead>
<tr>
<th>Item</th>
<th>( t )</th>
<th>( df )</th>
<th>Sig. (2-tailed)</th>
<th>Mean difference</th>
<th>95% Confidence Interval of the Difference</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>I could get information about work immediately by AR technology.</td>
<td>7.230</td>
<td>46</td>
<td>.000</td>
<td>.745</td>
<td>.54 to .95</td>
<td>3.74</td>
<td>.706</td>
</tr>
<tr>
<td>I could review peers’ works more clearly by using AR technology.</td>
<td>6.680</td>
<td>46</td>
<td>.000</td>
<td>.766</td>
<td>.54 to 1.00</td>
<td>3.77</td>
<td>.786</td>
</tr>
<tr>
<td>I could know the future application of peers’ works by using AR technology.</td>
<td>6.203</td>
<td>46</td>
<td>.000</td>
<td>.723</td>
<td>.49 to .96</td>
<td>3.72</td>
<td>.800</td>
</tr>
<tr>
<td>I could acquire detailed information about work to assess accurately by using AR technology.</td>
<td>3.545</td>
<td>46</td>
<td>.001</td>
<td>.468</td>
<td>.20 to .73</td>
<td>3.47</td>
<td>.905</td>
</tr>
<tr>
<td>I could know how work was created by using AR technology.</td>
<td>6.424</td>
<td>46</td>
<td>.000</td>
<td>.702</td>
<td>.48 to .92</td>
<td>3.70</td>
<td>.749</td>
</tr>
</tbody>
</table>

The results demonstrate that PA activities enable students to understand the teacher’s requests and engage their attention in learning. In addition, PA activities increase positive interaction through their discussions with peers. And regarding the attitudes toward the use of the system, most students believed that PA helps them to acquire more information about self-work and peers’ works and helps them to propose their viewpoints anytime anywhere by using the mobile service and AR technology.

For the progress of assessment grade, the average grade of peer assessment that had used MARPAS was 3.17, while the average grade of the 57 students who used pen-and-paper assessment was 2.43. To test the significance of this difference we have applied Mann-Whitney-Wilcoxon non-parametric test (since grades were not distributed normally), which showed that this difference is significant \( (U = 662, p = .02) \) in Table 5 and Figure 5.

<table>
<thead>
<tr>
<th>Method</th>
<th>Grade</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>MARPAS</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>Pen-and-paper</td>
<td>14</td>
<td>10</td>
</tr>
</tbody>
</table>

Through the experiment between pen-and-paper assessment and MARPAS assessment, the data collected in the survey demonstrated that students not only agreed about the positive effect of AR and mobile service, but also confirmed the usefulness of AR technology in learning. Students that used MARPAS had much better achievement, average \( M = 8.95 \) versus \( M = 5.79 \) points, (see Figure 5). This difference is also significant (test statistics \( U = 426, \) significance \( p = .00 \)). However, this difference can be attributed to the fact that PA activities
enabled students to understand the teacher’s requests and increase positive interaction through their discussions with peers. Most students believed that PA helped them to acquire more information about self-work and peers’ works and helped them to propose their viewpoints anytime anywhere by using the mobile service and AR technology. In addition, the qualitative analysis revealed that most participants thought that the system provided high autonomy and good visual effects. Importantly, the system helped students in acquiring rich and proper information while reviewing work, interacting with peers, and receiving assessment results. Moreover, use of the mobile service enabled students to propose their viewpoints anytime anywhere. The approach also eliminated the limitation of time, space, and devices.

By contrast, most students expressed that it was inconvenient to interact with peers in pen-and-paper peer assessment, and thus no student would like to use it. After the experiment, these students continued using MARPAS to interact with peers. Students that assessed their homework traditionally could collaborate, but they were not encouraged to do it and usually they worked independently, even on complex designs.

For the qualitative analysis, the feedback from the interviews indicated the following opinions about the use of MARPAS.

M01: “I can get the procedure and images of cooker on the work immediately! It’s so cool. I like this way of demonstration!”
M02: “There were no classmates beside when I assessed works by using this system. There was no pressure at all. And blind assessment could make me tell the truth.”
M07: “I am able to operate this system at home or at school. Even on the way to school in traffic, I can do it, too. This is very convenient. This convenience motivates me to give opinions to other classmates, and I hope to receive suggestions from classmates, too.”
M09: “Because of my inarticulateness, I cannot explain my works well. Now by using this system, I could prepare my illustration information and combine it with my works in advance. This is great to me. This will make me feel more confident about my works.”
M03: “I am not good at memory, and I forgot things quickly. The information showed on the works can be read repeatedly by me. So I won’t forget the detail of the works. This would make me do the correct assessment.”
M04: “The virtual information about the work can be procedure recorded video and even 3D digital model. It showed me things I have never seen before and things I would not see in physical work. This might affect my assessment.”

The above examples clearly show that AR technology did assist students during the assessment process.

Analysis of the interviews revealed that most participants credited the system with five advantages: high autonomy, good visual effects, prompt responses and rich assessment information from all assessors, convenient content management, and flexibility in using the system anytime anywhere.

Teachers provided the following feedback on the use of MARPAS:
T01: This system can assist teachers to arrange assessment activities more flexibly and make students more attentive to presentation, interaction, and feedback in the assessment process.
T02: For teachers, we can know how students review peers’ work from the AR situations and give more feedback to assessors and assessees. Moreover, teachers can create integrated AR situations for further applications.

**Conclusion**

This paper presents the novel system of an enhanced PA system complemented by the use of smart and mobile devices. Integrating the AR technology overcame some of the limitations of conventional PA systems, such as the implementation method, excessively high costs, and substantial errors. In this system, the AR technology enables students to observe how their peers completed their works by displaying videos of the cooking process over the completed dishes. During the assessment, students can determine whether their peers followed the instructions correctly by comparing the performance against the assessment criteria. By doing so, students can discover their own inadequacies or learn from other people’s methods. In addition, the system provides each student stable and convenient information and digital content based on environmental parameters or the identification of particular objects. Thus, students can learn while engaging in activities based on which their performance is assessed. Students can obtain appropriate learning information by using mobile devices to
photograph and identify target objects at appropriate moments and particular locations. The novel framework developed in this study, in which the AR technology was integrated, enables students to use various methods to observe the cooking processes and completed works of their peers. Simultaneously, the students can receive real-time feedback and recommendations regarding their own works. Hence, the barrier resulting from conflicting opinions between students can be eliminated, and students’ understanding of each other’s opinions can be enhanced. Thus, the accuracy of the results of PA can be improved.

Students who used MARPAS were encouraged to use its collaborative work feature. All students worked independently on their works. More complex cooking skill required students to know and combine many isometric exercises. Students who were using MARPAS were asked to do these exercises collaboratively, in groups of five students. Students usually communicated only with other members of their group, and had asked questions that other group members answered.

From the perspective of cognitive psychology, showing assessment criteria and feedback in real-time situations during assessment enables students to develop strong impressions of the feedback. Therefore, the students are highly capable of incorporating the feedback into their future work to achieve improvement and growth. The novel AR-integrated framework used in this experiment is almost complete. In instructional view, this system integrated the AR technology to overcome some of the limitations of conventional PA systems, such as subjective judgments of the evaluator, limited reliability, unsatisfactory economic performance, and low validity.

**Limitation and future plan**

“While it is important and useful to measure technical expectations, educators must beware of a blinkered vision that limits their ability to recognise and assess the expressive qualities that are enabled when technical skills are in place,” according to DeLuca and Bolden (2014). In fact, the National Research Council advocated in a recent report that moving to assessments of the next standard of science must be led by classroom-based assessments rather than by this complex endeavour with large-scale assessments (Pellegrino et al., 2014). The mobile AR technology is particularly suited to pilot this new approach, given their intensive efforts in implementing complex performance assessments. The information learned through competency-based assessments would then be used to support accountability determinations and, hopefully, better inform school improvement (e.g., Hargreaves & Braun, 2013). Thus, we will apply this system to some other courses, such as scene design, architecture, art design, and so on. Since different courses need different ways of PA, and this study did not undergo this research. Thus the dimensions of assessment should be studied as the next step, and we will use experimental research design to strengthen this study. And the experiment about how MARPAS assists teacher to manage the performance assessment and the course will be enforced.

Sometimes it’s hard to record the whole process of students’ performance by oneself. Students need a photographer for them if they are not working in a group. But, new technical devices such as Google Glass can solve this problem in the near future. This is convenient for all kind of students and courses.

**References**


