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 The articles should be original, unpublished, and not in consideration for publication elsewhere at the time of submission to Educational Technology & Society and three months thereafter.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

Editors
Kinshuk, Athabasca University, Canada; Demetrios G Sampson, University of Piraeus & ITI-CERTH, Greece; Nian-Shing Chen, National Sun Yat-sen University, Taiwan.

Editors' Advisors
Ashok Patel, CAL Research & Software Engineering Centre, UK; Reinhard Oppermann, Fraunhofer Institut Angewandte Informationstechnik, Germany

Editorial Assistant
Barbara Adamski, Athabasca University, Canada; Natalia Spyropoulou, University of Piraeus & Centre for Research and Technology Hellas, Greece

Associate editors
Vladimir A Fomichov, K. E. Tsiolkovsky Russian State Tech Univ, Russia; Olga S Fomicheva, Studio "Culture, Ecology, and Foreign Languages", Russia; Piet Koomers, University of Twente, The Netherlands; Chul-Hwan Lee, Inchon National University of Education, Korea; Brent Muirhead, University of Phoenix Online, USA; Erkki Sutinen, University of Joensuu, Finland; Vladimir Uskov, Bradley University, USA.

Assistant Editors
Yuan-Hsuan (Karen) Lee, National Chiao Tung University, Taiwan; Wei-Chieh Fang, National Sun Yat-sen University, Taiwan.

Advisory board
Ignacio Aedo, Universidad Carlos III de Madrid, Spain; Mohamed Ally, Athabasca University, Canada; Luis Anido-Rifón, 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 Harley, Leeds University, UK; J R Isaac, National Institute of Information Technology, India; Mohamed Jemni, University of Tunis, Tunisia; Mike Joy, University of Warwick, United Kingdom; Athanasia 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; Ruddy Lelouche, Universite Laval, Canada; Tzu-Chien Liu, National Central University, Taiwan; Roy McGreal, Athabasca University, Canada; David Merrill, Brigham Young University - Hawaii, USA; Marcelo Milrad, Växjö 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; Slavu Stoyanov, Open University, The Netherlands; Timothy Teo, Nanyang Technological University, Singapore; Chin-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.

Executive peer-reviewers
http://www.ifets.info/
Supporting Organizations
Centre for Research and Technology Hellas, Greece
Athabasca University, Canada

Subscription Prices and Ordering Information
For subscription information, please contact the editors at kinshuk@ieee.org.

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 kinshuk@ieee.org.

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.
All submissions should be in electronic form. The editors will acknowledge the receipt of submission as soon as possible.
The preferred formats for submission are Word document and RTF, but editors will try their best for other formats too. For figures, GIF and JPEG (JPG) are the preferred formats. Authors must supply separate figures in one of these formats besides embedding in text.
Please provide following details with each submission: • Author(s) full name(s) including title(s), • Name of corresponding author, • Job title(s), • Organisation(s), • Full contact details of ALL authors including email address, postal address, telephone and fax numbers.
The submissions should be uploaded at http://www.ifets.info/ets_journal/upload.php. In case of difficulties, please contact kinshuk@ieee.org (Subject: Submission for Educational Technology & Society journal).
## Special Issue Articles

**Guest Editorial: Overcoming the Technological Hurdles Facing Virtual Worlds in Education: The Road to Widespread Deployment**  
*Leonel Morgado, Baltasar Fernández Manjón and Christian Gütl*

Barriers and Enablers to the Use of Virtual Worlds in Higher Education: An Exploration of Educator Perceptions, Attitudes and Experiences  
*Sue Gregory, Sheila Scutter, Lisa Jacka, Marcus McDonald, Helen Farley and Chris Newman*

Coming Down to Earth: Helping Teachers Use 3D Virtual Worlds in Across-Spaces Learning Situations  
*Juan A. Muñoz-Cristóbal, Luis P. Prieto, Juan I. Asensio-Pérez, Alejandra Martínez-Monés, Iván M. Jorrín-Abellán and Yannis Dimitriadis*

Toward Educational Virtual Worlds: Should Identity Federation Be a Concern?  
*Gonçalo Cruz, António Costa, Paulo Martins, Ramiro Gonçalves and João Barroso*

Technical Problems Experienced in the Transformation of Virtual Worlds into an Education Environment and Coping Strategies  
*Murat Coban, Turkan Karakus, Asiye Karaman, Fatma Gunay and Yüksel Goktas*

Distributed Pervasive Worlds: The Case of Exergames  
*Teemu H. Laine and Carolina Islas Sedano*

## Full Length Articles

What is the Relationship between Technology and Mathematics Teaching Anxiety?  
*Enver Tatar, Yılmaz Zengin and Türkan Berrin Kağızmanlı*

Validating the Learning Cycle Models of Business Simulation Games via Student Perceived Gains in Skills and Knowledge  
*Yu-Hui Tao, C. Rosa Yeh and Kung Chin Hung*

A Learning Cycle Approach to Developing Educational Computer Game for Improving Students’ Learning and Awareness in Electric Energy Consumption and Conservation  
*Ugyen Dorji, Patcharin Panjaburee and Niwat Srisawasdi*

The Impact of Digital Mobile Devices in Higher Education  
*M.ª Luisa Sevillano-García and Esteban Vázquez-Cano*

Aptitude-treatment Interactions during Creativity Training in E-Learning: How Meaning-making, Self-regulation, and Knowledge Management Influence Creativity  
*Yu-chu Yeh and Chun Fu Lin*

Development and Use of an Adaptive Learning Environment to Research Online Study Behaviour  
*Anna Helga Jonsdottir, Audbjorg Jakobsdottir and Gunnar Stefansson*

Integrating learning services in the Cloud: An approach that benefits both systems and learning  
*Gustavo Gutiérrez-Carréon, Thanasis Daradoumis and Josep Jorba*

Exploring the Profiles and Interplays of Pre-service and In-service Teachers’ Technological Pedagogical Content Knowledge (TPACK) in China  
*Yan Dong, Ching Sing Chai, Guo-Yuan Sang, Joyce Hwee Ling Koh and Chin-Chung Tsai*

Blending Face-to-Face Higher Education with Web-Based Lectures: Comparing Different Didactical Application Scenarios  
*Hannelore Montrieux, Sandra Vangestel, Annelies Raes, Paul Matthys, Tammy Schellens*

A Pilot Meta-Analysis of Computer-Based Scaffolding in STEM Education  
*Brian R. Belland, Andrew E. Walker, Megan Whitney Olsen and Heather Leary*
A Two-Tier Test-based Approach to Improving Students’ Computer-Programming Skills in a Web-Based Learning Environment
  Tzu-Chi Yang, Gwo-Jen Hwang, Stephen J. H. Yang and Gwo-Haur Hwang  

The Exploration of Elementary School Teachers’ Internet Self-Efficacy and Information Commitments: A Study in Taiwan
  Ying-Tien Wu and Li-Jen Wang  

Social Presence and Interaction in Learning Environments: The Effect on Student Success
  Ines Kožuh, Zoran Jeremić, Andrej Sarjaš, Julija Lapuh Bele, Vladan Devedžić and Matjaž Debevc  

Exploring the Relationship between Sanctioned and Unsanctioned Laptop Use in a 1:1 Classroom
  Martin Tallvid, Johan Lundin, Lars Svensson and Berner Lindström  

The Effectiveness of Adopting E-Readers to Facilitate EFL Students’ Process-Based Academic Writing
  Hui-Chun Hung and Shelley Shwu-Ching Young  

Effects of Worked Examples Using Manipulatives on Fifth Graders’ Learning Performance and Attitude toward Mathematics
  Chun-Yi Lee and Ming-Jang Chen  

Learning Faults Detection by AIS Techniques in CSCL Environments
  Amina Zedadra and Yacine Lafifi  

Role of Parents and Annotation Sharing in Children’s Learning Behavior and Achievement Using E-Readers
  Wu-Yuin Hwang, Yi-Fan Liu, Hon-Ren Chen, Jian-Wan Huang and Jin-Yi Li  

Mobile Enhanced Learning in a South African Context
  Mmaki Jantjies and Mike Joy  

Measuring Knowledge Elaboration Based on a Computer-Assisted Knowledge Map Analytical Approach to Collaborative Learning
  Lanqin Zheng, Ronghuai Huang, Gwo-Jen Hwang and Kaicheng Yang  

Does Gender Influence Emotions Resulting from Positive Applause Feedback in Self-Assessment Testing?
  Evidence from Neuroscience
  Chia-Ju Liu, Chin-Fei Huang, Ming-Chi Liu, Yu-Cheng Chien, Chia-Hung Lai and Yueh-Min Huang
Guest Editorial: Overcoming the Technological Hurdles Facing Virtual Worlds in Education: The Road to Widespread Deployment

Leonel Morgado¹², Baltasar Fernández Manjón³ and Christian Gütl⁴⁵

¹INEC TEC, Porto, Portugal // ²Universidade Aberta, Lisbon, Portugal // ³Universidad Complutense de Madrid, Madrid, Spain // ⁴Graz University of Technology, Graz, Austria // ⁵Curtin University of Technology, Perth, Western Australia // leonel.morgado@uab.pt // balta@fdi.ucm.es // cgutl@iicm.edu

Overwhelmingly, most current studies on the field of educational use of virtual worlds focus on educational theory and practice (Hew & Cheung, 2010). While we see this as natural in a field centred on pedagogical practice, it is also a shortcoming in the large body of literature developed on this subject in recent years. Education-centred studies tend to adopt a static view of technology, as something that has to be employed in its current state or something that was merely developed with the specific intent of fulfilling a predefined purpose (Duncan et al., 2012). Consequently, the technological perspective of such studies typically limits itself to identifying limitations/restrictions of specific virtual worlds – not of the concept of virtual worlds in general.

This is in stark contrast with the technological research viewpoint: from it, virtual worlds are artefacts. That is, they are embodiments of development processes, concrete renderings of knowledge (Hevner, 2007). By coming into existence, they generate new technological knowledge, and enable new educational processes. Hence, they are not just static facts, to be dealt with “as is.”

This special issue was launched following this technological perspective, and called for contributions towards an updated understanding of the technological challenges that virtual worlds face when applied to education and training, and towards an identification of research directions to overcome them. As a starting point, three fields were suggested as open for contributions, albeit contributors could also suggest others:

- making the virtual worlds technology available to educational actors;
- content production in virtual worlds to support advanced interactions beyond plain dialog;
- large scale deployment of virtual worlds and integration with current information systems.

The paper by Gregory et al. helps set a perspective on the challenges this technology faces, by aiming to provide an overview of barriers and enablers from the educational actors’ side. They attempt to determine why virtual worlds not only didn’t become mainstream but even lost some popularity, based on a survey of over 200 academics, with results examined in view of their prior, current, and expected use of virtual worlds.

Muñoz-Cristóbal et al.’s paper provides a feasibility example. This paper addresses the difficulties teachers face in order to create and share their designs across virtual worlds, as well as the (lack) of integration between virtual worlds and other educational technologies. They present a system focusing on learning situations across different platforms, including Web-based, augmented reality, and virtual Earth globes. The paper evaluates the system in the context of teacher-based deployment of learning activities across different spaces, providing data to support the adoption virtual worlds in everyday teaching and institutional practice.

Cruz et al.’s paper steps back from teaching practice and takes an organizational perspective. They reflect on why and how identity federation may be required to overcome barriers faced by institutions wishing to use virtual worlds. Their set of scenarios present this issue across various organizational dimensions of the educational process, and expose it via concerns such as identity, traceability, privacy, accountability, and interoperability. Their presentation of this issue is completed with pathways towards operational solutions.

Coban et al.’s paper contributes with a plethora of everyday practice and organizational details – as well as impacts on pedagogical issues – providing a better understanding of how technology must change in order for its widespread use to become a reality. Their data stem from a welcome one-year longitudinal observation of two cases, using two different virtual worlds, from the perspectives of designers, practitioners, and administrators.

This issue concludes with a final paper, by Laine & Sedano, who bring to fore the issue of virtual worlds evolution in the light of the current trend towards pervasive technology. Their case study explores the concept, possibilities, and challenges of combining a virtual game with sensor-enhanced players’ activity in the physical world. Through their
study, valuable steps are made towards design, development, and evaluation of virtual worlds that combine virtual and physical elements.

We consider that this special issue contributes to provide an updated and alternative point of view from the technological perspective about some aspects that could help to overcome some of the identified issues with the actual deployment and use of virtual worlds.

Finally, we would like to thank all the reviewers for their great work in selecting and improving the papers with all their insightful comments.

References


Barriers and Enablers to the Use of Virtual Worlds in Higher Education: An Exploration of Educator Perceptions, Attitudes and Experiences

Sue Gregory*, Sheila Scutter, Lisa Jacka, Marcus McDonald, Helen Farley and Chris Newman

University of New England, Australia // University of South Australia, Australia // Southern Cross University, Australia // Royal Melbourne Institute of Technology University, Australia // University of Southern Queensland, Australia // Queensland University of Technology, Australia // sue.gregory@une.edu.au // Sheila.Scutter@unisa.edu.au // lisa.jacka@scu.edu.au // marcus.mcdonald@rmit.edu.au // Helen.Farley@usq.edu.au // christine.newman@qut.edu.au

*Corresponding author

ABSTRACT

Three-dimensional (3D) virtual worlds have been used for more than a decade in higher education for teaching and learning. Since the 1980s, academics began using virtual worlds as an exciting and innovative new technology to provide their students with new learning experiences that were difficult to provide any other way. But since that time, virtual worlds have failed to maintain their popularity as learning spaces; many builds falling into disuse and many disappearing altogether. The aim of this article is not only to determine why virtual worlds have not become a mainstream teaching tool, but to ascertain why they have even failed to maintain their popularity. In order to do this, the research team surveyed over 200 academics about the barriers and enablers to the use and perceived affordances of virtual worlds in teaching and learning. These responses are examined in relation to academics’ past, present and future use, experience and knowledge of virtual world environments.

Keywords
Virtual worlds, Educator attitudes, Teaching and learning

Introduction

Interest in the use of virtual worlds in teaching has been maintained since their first wide scale use in the 1980s (Warburton, 2009). However, despite the persistence of interest, and with the development of many teaching facilities in virtual worlds, teaching in these environments has not become mainstream, and the numbers of educators using this environment for teaching is in fact decreasing. This is evidenced by the number of underutilized and disused builds that are seen in virtual worlds. When exploring virtual worlds such as Second Life, and looking for interesting or useful education sites to visit, another avatar may not even be encountered when visiting these spaces. What has become of the islands which have been reduced to virtual dust when their rent has lapsed or of the privately hosted virtual worlds sitting idle on their own dusty corner of a server? Why has so much effort gone into the development of these sites for them only to be discarded? After the considerable investment of resources in virtual world learning spaces, it might be expected that the use of virtual worlds would have reached the “plateau of productivity” on the Gartner Hype Cycle (Linden & Fell, 2003). However, recent analysis suggests they are still in the “trough of disillusionment” from which point, many new technologies fade into disuse unless maintained and their use progressed by a dedicated group of users and innovators (Fenn, 2008).

Virtual worlds are able to provide a diverse and relatively inexpensive environment compared to bricks and mortar, suitable for authentic learning experiences, potentially removing the tyranny of distance for students studying away from campus (Ritzema & Harris, 2008). They accommodate a range of learning styles (Bonk & Zhang, 2006) and provide risk-free access (Bronack, Sanders et al. 2008) to dangerous, complex or expensive environments (Monahan, Ullberg & Harvey, 2009). When teaching in virtual worlds, Steve Bronack and his colleagues describe the utilization of presence pedagogy, grounded in social constructivist theory, in which students and instructors become part of a community of practice, where all have the potential to be learners, teachers, peers and/or experts (Bronack, Sanders et al. 2008). This encourages reflective learning and engagement in the process (Boulou, Hetherington et al. 2007). If the pedagogy is sound and the initial investment of time and resources in developing the learning environment has been provided, why then are these spaces under-utilized or abandoned, and further, why have they not been mainstreamed? This research aims to identify issues that influence academics in their decisions about whether or not to use virtual worlds in their teaching, whether they are already using them or contemplating using them in the future. This article presents the results from a survey investigating educators’ experiences and plans for future use of virtual worlds in teaching, and the issues that influenced these decisions. Two of the researchers have been involved in the development and teaching of higher education courses in virtual worlds since 2007.
Literature review

In 2007, the Horizon Report predicted a two to three year time frame for the adoption of virtual worlds in education (The New Media Consortium, 2007). Similarly, the subsequent Australia New Zealand Horizon Report (Johnson, Levine, & Smith, 2008, p. 6) enthusiastically identified virtual worlds as “spaces for truly immersive forms of learning and for a level of collaboration that is erasing traditional boundaries and borders rapidly.” In 2008 and again in 2009, the Gartner hype cycle positioned virtual worlds in higher education at the peak of its hype cycle (Lowendahl et al., 2008) and the information technology research and advisory company famously predicted that 80% of Internet users would have an avatar in a virtual world by the end of 2011 (Stamford, 2007). Many universities across the globe set up spaces and experimented with ways to leverage the learning opportunities virtual worlds afforded.

E-learning experts were hopeful for the future of virtual worlds. Chea (2007, p. 205) stated that the success of these environments was almost inevitable due to the post-modern focus of our society on “irony, ambiguity, fragmentation, plurality and globalization” and the use of these environments in education flowed on from this as they encouraged a constructivist and immersive paradigm for student engagement as educators moved away from a didactic approach to teaching. However, in recent years, not only have the interest and activity in virtual worlds not been reflected by their widespread adoption, their popularity among educators has declined as evidenced by the underutilization of many existing builds. The Gartner Hype Cycle attempts to describe the typical pattern in the adoption of new technologies (see Figure 1). Hype cycles are used across many disciplines attempting to predict the changing use or adoption of a given idea or product. While they are based heavily in theory, they may provide us a useful guild to shape future efforts (van Lente, Spitters & Peine, 2013). After an initial “technology trigger” when a new technology appears to have potential for significant improvements, a “peak of inflated expectations” occurs, where initial success stories suggest that the application of the new technology will be broad and successful. The “trough of disillusionment” follows, when expectations are not achieved and some projects fail. If the technology continues to be used, the “slope of enlightenment” may follow, during which the technology is refined and its place in the existing world is better understood. Eventually, a “plateau of productivity” may be achieved, in which the technology becomes embedded in mainstream activity. In 2010, Gartner positioned virtual worlds as sliding into the “trough of disillusionment” and they were still there in 2013 (Lowendahl, 2013). Many authors suggest that one cause of the decline is the prohibitive cost in both time and money in a difficult economy. There is considerable time involved setting up and maintaining a place in a virtual world: As with “real” worlds, upkeep is required. In order to put the time and effort into developing virtual teaching spaces, academics need to be convinced that the educational outcomes are improved, at least for a majority of students.

Much research into virtual worlds in education has focused on the potential for improved learning. For example, Standen, Brown and Comby (2001) demonstrated that virtual worlds could be effective in teaching living skills to people with an intellectual disability due to the simulated real world environment. In a review of the literature, Eschenbrenner, Fui-Hoon Nah, and Siau (2008) describe the benefits and potential limitations of teaching in the virtual world, but there is little available in terms of evidence of improved learning outcomes. Although Wiecha, Heyden, Sterntahl and Merialdi (2010) demonstrated that learning outcomes in a small group of physicians undertaking Continuing Medical Education were positive and that participants liked the virtual world for learning, there was no comparison with traditional learning approaches. Triola (2006) established that there were no differences in learning outcomes between using Standardized Patients (SP) and Virtual Patients (VP) in a medical education program. Triola goes on to suggest that a significant advantage of the VP was the cost, standardization, reusability and access. However, most of the literature pertaining to virtual worlds is descriptive in nature, and although positive feedback from both educators and students is reported, more substantial evidence of outcomes may be required to convince academics to invest the time and resources required.

![Figure 1. Gartner Hype Cycle](image-url)
To successfully teach in virtual worlds, academics must not only learn the skills to use the space but also to understand the pedagogical affordances involved (Smith-Robbins, 2011). Institutional technical requirements are often difficult to implement and maintain, and are often cited as a significant barrier to uptake (Dudney & Ramsay, 2009). The virtual world software is continually being developed, may need updating and therefore can be unreliable or in need of significant amounts of maintenance. Insufficient bandwidth can be problematic, especially if several computers are sharing a network. Also, because digital assets are not stored locally, they are streamed and rendered in real time, further taxing Internet bandwidth. Other reasons suggested for the disinterest is the lack of support to educators in terms of technical and pedagogical support or provision of additional time to develop virtual world lessons (Young, 2010; Smith-Robbins, 2011). Many are worried about the stability of providers: several virtual worlds have been discontinued leaving users in the lurch (Young, 2010). Yoon and George (2013) explored reasons why organizations have not adopted virtual worlds more widely. They developed a model based on the Technology-Organization-Environment Framework. This research differs from that of Yoon and George as we are interested in why educators have invested time and money into the development of virtual world spaces and activities and are no longer utilizing it. Why have these builds been abandoned or disappeared altogether? We also want to know why other educators have not taken up the challenge of using virtual worlds as a research, teaching and learning tool when most Australian higher education institutions have used, or are using, a virtual world at their institutions (Gregory et al., 2012).

In order to understand why virtual worlds are not being adopted by educators at the rate anticipated, it is useful to understand the performance pressures on higher education institutions. Yoon and George (2013) state that the organizational adoption of technology innovations can be significantly influenced by institutional pressure that helps them achieve organizational legitimacy. Therefore, an institution will adopt an innovative technology if other institutions are also doing so, a kind of eminence-based practice. Furthermore, they suggest looking to the adoption of the Internet as being similar to the adoption of virtual worlds with “relative advantage” or “the degree of perceived benefit” being a key driver in its adoption (p. 776). In terms of the adoption of ICTs in education, a number of studies have utilised Davis’ (1986) Technology Adoption Model, the premise being that the users’ perceived usefulness and perceived ease of use might influence user acceptance. This model has been adopted in an education setting specifically with an investigation into school teachers’ use of technology (Cox, Preston & Cox, 1999) and with pre-service teachers (Sime & Priestley, 2005; Gill & Dalgarno, 2010; Teo, Lee & Chai, 2008). It is our experience that this applies directly to some educators using virtual worlds. This may be especially the case where the educator is not the initiator of the virtual world project but is perhaps influenced by students’ experiences either through direct feedback or the institution’s student experience survey system (Salomon, 2010). McDonald, Ryan and colleagues (2012) discuss how student-reported technical difficulties were more likely to be due to user error. Selwyn (2009) describes the current generation’s use, and perhaps we can infer too, understanding of technology, as “unspectacular”. It would seem that it is not always a case of “build it and they will come.” Staff may choose not to use educational tools that students find difficult and which could result in complaints, poor scoring on student experience surveys and no significant outcomes. In addition, virtual worlds are just one more technology vying for the attention of students and educators (Smith-Robbins, 2011; Essid, 2012). The current trend in technology adoption is towards more available and ubiquitous mobile devices and social media. Until virtual worlds become more intuitive to use and can be accessed ubiquitously, they will lose the battle for prominence (Salmon, Nie, & Edirisingha, 2010).

There are a number of issues to overcome before virtual worlds become a mainstream teaching tool to be used by educators. A scoping study undertaken in Australia and New Zealand reviewed and analysed the use of 3D virtual worlds in teaching and learning in higher education (see Dalgarno, Lee, Carlson, Gregory, & Tynan, 2011a; 2011b; 2010; Dalgarno, Gregory, Carlson, Lee, & Tynan, 2013; Lee, Dalgarno, Gregory, Carlson, & Tynan, 2013). A number of problems were identified in relation to integrating virtual worlds into teaching and learning. Various categories of challenges were formulated from the feedback collected from educators using virtual worlds, including “[lack of] technology, support, funding and time, usability and familiarity, equity and ethics, inherent limitations of virtual worlds, acceptance of virtual worlds, and management and planning” (Dalgarno et al., 2013, p. 10).

**Methodology**

In order to discern why virtual worlds did not emerge as a mainstream teaching tool and in fact are declining in popularity with educators, it was decided to examine the factors that influenced the adoption and continued use of virtual worlds by educators. The research team, drawing from their own experiences with virtual worlds, envisaged a number of potential factors that could impact their adoption. The research team comes from a range of disciplines including the humanities, education and the health sciences. Some of the teams are teaching academics, some are in teaching and learning support and some are research-focused. This range of roles and perspectives gave rise to a broad experience of using virtual worlds across many contexts, allowing for the identification of a broad range of issues including lack of technical support, lack of expertise, insufficient funding, and student attitudes among others.

The University of New England’s Human Research Ethics Committee provided ethics approval for this study. A survey was designed by the authors and distributed in June 2013. The request to complete the survey was sent out to the researchers’ various networks from within their institution and working in virtual worlds education. The survey was distributed to members of the
Australian and New Zealand Virtual Worlds Working Group, virtual world users of LinkedIn, Facebook, various virtual world lists, members of the researchers’ institutions and virtual world user groups of other organisations.

The survey questions were created with the primary intention of eliciting responses in relation to why educators have persisted with or ceased with the use of virtual worlds. More specifically one of the questions focused on those that are not teaching in virtual worlds. Respondents were asked to choose from seventeen pre-designed answers or offering their own. The pre-designed responses were created from the range of experiences reported by the researchers themselves and from the review of the available literature on virtual worlds in higher education. The potential issues that influenced the design of the question responses had been identified as relating to institutional support, student support, technical support and access.

A subset to the question about why educators were not using virtual worlds was whether they perceived value in the use of virtual worlds. In order to identify the perceptions of the value of virtual worlds in education, respondents were asked to rate the importance of five identified learning benefits in relation to their discipline. The learning benefits were chosen based on the researchers’ experience and a review of the literature.

A series of demographic questions were asked to gather data about the platforms used, class types and size, respondent’s age and institution. Respondents were asked to indicate the time period in which they had been involved with teaching with virtual worlds. The choices were - now (to mean the date at which they undertook the survey), in the past or intended to in the future. They were also asked to indicate in which years they had used virtual worlds with specific reference to four time frames (2000 and earlier, 2001-2005, 2006-2010, 2011 and after).

**Results/analysis**

The survey was completed by 223 respondents. Responses were received from 134 institutions in 28 countries, including 38 from Australia, 37 from the United States of America, 12 from the United Kingdom, seven from Canada, three from New Zealand and Finland and the rest were made up of individual users from different countries, particularly in Europe and South America.

The largest age group of respondents was in the 46–55 years category (38%). The other age groups were represented in smaller numbers; 25 or under (4%), 26–35 (10%), 36–45 (24%), 56–65 (20%) and over 65 (5%). The 46–55 age group were the ones most likely to have stopped using virtual worlds after initially using them with 7% (15/223) indicating they had used virtual worlds in the past but no longer use them compared to the other age groups; 25 or under (0%), 26–35 (0.5%, 1/223), 36–45 (3%, 7/223), 56–65 (2%, 4/223) and over 65 (0%). Across all age groups the number of users who had previously used virtual worlds but no longer used them was 18% (36/204). Although the largest respondents were from the 46–55 age category, they were also the largest group of users who have discontinued using virtual worlds.

Twelve distinct groups of users emerged according to their past, present and future use of virtual worlds (see Table 1). The majority of respondents (52%, 110/204) were not currently using virtual worlds for teaching and 36% (74/204) had never used virtual worlds for teaching. Of the 74 who had not used virtual worlds 60% (44/74) said they might use virtual worlds in the future, 23% (17/74) said they would be using it in the future and 17% (13/74) indicated that they had no plans to use virtual worlds. The respondents who currently use virtual worlds (94/204) indicated that 84% (79/94) had used them in the past and 90% (85/94) would use them in the future.

**Table 1. Past, present and future teaching in virtual worlds**

<table>
<thead>
<tr>
<th>Past</th>
<th>Present</th>
<th>Future</th>
<th>Number of responses</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>No</td>
<td>No</td>
<td>13</td>
<td>6%</td>
</tr>
<tr>
<td>No</td>
<td>No</td>
<td>Maybe</td>
<td>44</td>
<td>22%</td>
</tr>
<tr>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>17</td>
<td>8%</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>Maybe</td>
<td>3</td>
<td>1%</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>12</td>
<td>6%</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>4</td>
<td>2%</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>Maybe</td>
<td>18</td>
<td>9%</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>14</td>
<td>7%</td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>Maybe</td>
<td>6</td>
<td>3%</td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>73</td>
<td>36%</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>204</td>
<td></td>
</tr>
</tbody>
</table>
Of those who had used virtual worlds, 71 (47%) indicated that Second Life was their preferred platform compared with a range of other platforms including: OpenSim (23%), Jibe or other Unity 3D virtual words (7%), with 3% indicating Reaction Grid or Kitely. Thirteen percent indicated they used a variety of “in house” or “bespoke” virtual worlds.

Responses to the survey showed a range of experience in using virtual worlds. Four percent of individuals (5/114) indicated that they had been using virtual worlds prior to 2001. Of those who are currently using virtual worlds, 76% (60/79) first used them in the period between 2006 and 2010. Equally significant is that the same period (2006–2010) is the time in which the majority (62%, 22/35) of those who are no longer using virtual worlds first used them. These time frames are indicative of Gartner’s Hype Cycle, which indicated in 2009 that virtual worlds were at their peak and were heading towards the trough of disillusionment.

Why educators are not teaching in virtual worlds

The responses to the question asking for reasons why educators are not teaching in virtual worlds (see Table 2) were clustered into four groups according to those that related to:

- Technological issues (T)
- Potential student difficulties (S)
- Institutional issues (I)
- Personal perceptions (P)

Table 2 illustrates the responses to these questions grouped by those currently using virtual worlds and those who are not. Table 3 unpacks the data by looking solely at the group who are not currently teaching in virtual worlds.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Response numbers</th>
<th>Those currently using virtual worlds (17)</th>
<th>Those not currently using virtual worlds (95)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>My institution doesn’t provide adequate technology to use virtual worlds</td>
<td>30%</td>
<td>44%</td>
</tr>
<tr>
<td>I</td>
<td>My institution doesn’t provide funding to use virtual worlds</td>
<td>30%</td>
<td>40%</td>
</tr>
<tr>
<td>I</td>
<td>My institution doesn’t provide teaching support to use virtual worlds</td>
<td>36%</td>
<td>29%</td>
</tr>
<tr>
<td>I</td>
<td>My institution doesn’t provide technical support to use virtual worlds</td>
<td>54%</td>
<td>37%</td>
</tr>
<tr>
<td>P</td>
<td>I don’t have the time</td>
<td>6%</td>
<td>29%</td>
</tr>
<tr>
<td>P</td>
<td>It’s all just too hard</td>
<td>6%</td>
<td>11%</td>
</tr>
<tr>
<td>P</td>
<td>I feel uncomfortable in the virtual world environment</td>
<td>0%</td>
<td>7%</td>
</tr>
<tr>
<td>P</td>
<td>I don’t have the computing skills to use virtual worlds in teaching</td>
<td>6%</td>
<td>17%</td>
</tr>
<tr>
<td>P</td>
<td>My colleagues don’t think it is a good idea</td>
<td>12%</td>
<td>5%</td>
</tr>
<tr>
<td>P</td>
<td>No-one else I know is using them</td>
<td>18%</td>
<td>20%</td>
</tr>
<tr>
<td>P</td>
<td>My students gave poor feedback</td>
<td>12%</td>
<td>5%</td>
</tr>
<tr>
<td>P</td>
<td>Virtual worlds are just a game and not suitable for use in teaching and learning at a tertiary institution</td>
<td>0%</td>
<td>2%</td>
</tr>
<tr>
<td>P</td>
<td>My classes are going very well as they are</td>
<td>6%</td>
<td>4%</td>
</tr>
<tr>
<td>P</td>
<td>I’ve heard they are a poor educational tool</td>
<td>0%</td>
<td>1%</td>
</tr>
<tr>
<td>S</td>
<td>My classes are too big</td>
<td>6%</td>
<td>4%</td>
</tr>
<tr>
<td>S</td>
<td>I don’t want my students exposed to the kind of material you can come across in virtual worlds</td>
<td>18%</td>
<td>2%</td>
</tr>
<tr>
<td>T</td>
<td>The virtual world I use is too unreliable</td>
<td>6%</td>
<td>9%</td>
</tr>
</tbody>
</table>

A further breakdown of the respondents focusing on the ones who are not currently using virtual worlds in teaching shows some differences in their reasons. Users who had used virtual worlds in the past but will not use them in the future cited the unreliability of the virtual world as a reason, as did the group who had never used virtual worlds. More significant was the choice of “feeling uncomfortable in the virtual world environment” as those who would not use virtual worlds in the future chose this, yet those who may or those who will, did not choose it at all. Table 3 shows the responses to the question: “What are your reasons for not teaching in virtual worlds?” from the group of respondents who are not currently teaching in virtual worlds. Each of the columns indicates whether the respondents had used virtual worlds in the past and whether they were considering using virtual worlds in the future.
Table 3. Reasons for not teaching in virtual worlds

<table>
<thead>
<tr>
<th>Question: What are your reasons for not teaching in virtual worlds?</th>
<th>Have used virtual world in the past</th>
<th>Planning to use virtual worlds in the future</th>
<th>Response to question</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
<td>Maybe</td>
<td>Yes</td>
</tr>
<tr>
<td>My institution doesn’t provide adequate technology to use virtual worlds</td>
<td>40%</td>
<td>45%</td>
<td>21%</td>
</tr>
<tr>
<td>My institution doesn’t provide funding to use virtual worlds</td>
<td>40%</td>
<td>45%</td>
<td>21%</td>
</tr>
<tr>
<td>My institution doesn’t provide teaching support to use virtual worlds</td>
<td>40%</td>
<td>38%</td>
<td>29%</td>
</tr>
<tr>
<td>My institution doesn’t provide technical support to use virtual worlds</td>
<td>50%</td>
<td>43%</td>
<td>14%</td>
</tr>
<tr>
<td>The virtual world I use is too unreliable</td>
<td>20%</td>
<td>5%</td>
<td>0%</td>
</tr>
<tr>
<td>I don’t have the time</td>
<td>30%</td>
<td>40%</td>
<td>7%</td>
</tr>
<tr>
<td>It’s all just too hard</td>
<td>20%</td>
<td>15%</td>
<td>0%</td>
</tr>
<tr>
<td>I feel uncomfortable in the virtual world environment</td>
<td>10%</td>
<td>10%</td>
<td>0%</td>
</tr>
<tr>
<td>I don’t have the computing skills to use virtual worlds in teaching</td>
<td>40%</td>
<td>25%</td>
<td>14%</td>
</tr>
<tr>
<td>My colleagues don’t think it is a good idea</td>
<td>20%</td>
<td>0%</td>
<td>7%</td>
</tr>
<tr>
<td>No-one else I know is using them</td>
<td>50%</td>
<td>27%</td>
<td>7%</td>
</tr>
<tr>
<td>My students gave poor feedback</td>
<td>0%</td>
<td>3%</td>
<td>0%</td>
</tr>
<tr>
<td>My classes are too big</td>
<td>0%</td>
<td>5%</td>
<td>0%</td>
</tr>
<tr>
<td>Virtual worlds are just a game and not suitable for use in teaching and learning at a tertiary institution</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>I don’t want my students exposed to the kind of material you can come across in virtual worlds</td>
<td>0%</td>
<td>0%</td>
<td>7%</td>
</tr>
<tr>
<td>My classes are going very well as they are</td>
<td>0%</td>
<td>8%</td>
<td>0%</td>
</tr>
<tr>
<td>I’ve heard they are a poor educational tool</td>
<td>0%</td>
<td>3%</td>
<td>0%</td>
</tr>
</tbody>
</table>

The role of virtual worlds in specific disciplines

The majority (82%) of all respondents stated that they did believe that virtual worlds have a role to play in their particular discipline. All of the current users agreed that virtual worlds have a role to play. The respondents who either felt that they did not have a role to play (4%) or were unsure (14%) were predominantly from the group of respondents who had never used virtual worlds either now or in the past (see Table 4). There was a direct link to the group who had never used virtual worlds but were considering using them, with the largest response to being unsure about whether the virtual world has a role to play in their discipline. This correlation might suggest that the barrier to use of virtual worlds in teaching is not only connected to the institutional support but also to whether the educator has been introduced to the possible benefits in terms of their specific disciplinary context. Table 4 connects the beliefs of educators that virtual worlds have a role to play in their discipline, with the actual use they have had with virtual worlds in their teaching.

Table 4. Comparison of use of virtual worlds in teaching and whether they have a role to play in the respondent’s discipline

<table>
<thead>
<tr>
<th>Use of virtual worlds in their teaching</th>
<th>Past</th>
<th>Present</th>
<th>Future</th>
<th>Yes</th>
<th>No</th>
<th>Unsure</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>No</td>
<td>No</td>
<td>30%</td>
<td>30%</td>
<td>40%</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>No</td>
<td>Maybe</td>
<td>55%</td>
<td>5%</td>
<td>40%</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>86%</td>
<td>0%</td>
<td>14%</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>Maybe</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>92%</td>
<td>8%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>0%</td>
<td>25%</td>
<td>75%</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>Maybe</td>
<td>88%</td>
<td>0%</td>
<td>12%</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>Maybe</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td></td>
</tr>
</tbody>
</table>
Respondents were also asked to rate five learning benefits of 3D immersive virtual worlds. Consistently, the majority of respondents rated the five learning benefits as very important or important (see Table 5). Interestingly, for this research, are the respondents who have never used virtual worlds for teaching. Despite having not used virtual worlds, they still rated the learning benefits as important. The consistency in the responses showing that educators believe that virtual worlds have a role to play in their discipline and that they have specific learning benefits of value to their discipline highlight that there are factors outside of pedagogical potential that influence the use of virtual worlds. Table 5 shows the response rates to the question that asked respondents to rate the importance of five identified learning benefits of virtual worlds for their discipline. The table compares data between three groups; (1) those who have never used virtual worlds, (2) those who have used virtual worlds but are not currently using them, and (3) those who currently use them.

<table>
<thead>
<tr>
<th>Learning Benefit 1 – They can assist learners in developing familiarity with a place and the objects within it</th>
<th>62 responses</th>
<th>35 responses</th>
<th>93 responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Important</td>
<td>16%</td>
<td>26%</td>
<td>42%</td>
</tr>
<tr>
<td>Important</td>
<td>58%</td>
<td>40%</td>
<td>43%</td>
</tr>
<tr>
<td>Neither</td>
<td>13%</td>
<td>14%</td>
<td>13%</td>
</tr>
<tr>
<td>Unimportant</td>
<td>6%</td>
<td>11%</td>
<td>1%</td>
</tr>
<tr>
<td>Very unimportant</td>
<td>6%</td>
<td>9%</td>
<td>1%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Learning Benefit 2 – They can be motivating and engaging to learners</th>
<th>63 responses</th>
<th>35 responses</th>
<th>93 responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Important</td>
<td>37%</td>
<td>46%</td>
<td>62%</td>
</tr>
<tr>
<td>Important</td>
<td>54%</td>
<td>43%</td>
<td>33%</td>
</tr>
<tr>
<td>Neither</td>
<td>6%</td>
<td>6%</td>
<td>3%</td>
</tr>
<tr>
<td>Unimportant</td>
<td>2%</td>
<td>3%</td>
<td>1%</td>
</tr>
<tr>
<td>Very unimportant</td>
<td>2%</td>
<td>3%</td>
<td>0%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Learning Benefit 3 – They can lead to improved transfer of learning to real situations</th>
<th>61 responses</th>
<th>35 responses</th>
<th>93 responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Important</td>
<td>28%</td>
<td>43%</td>
<td>65%</td>
</tr>
<tr>
<td>Important</td>
<td>67%</td>
<td>49%</td>
<td>32%</td>
</tr>
<tr>
<td>Neither</td>
<td>3%</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>Unimportant</td>
<td>2%</td>
<td>3%</td>
<td>0%</td>
</tr>
<tr>
<td>Very unimportant</td>
<td>0%</td>
<td>3%</td>
<td>0%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Learning Benefit 4 – They can enable more effective collaborative learning</th>
<th>61 responses</th>
<th>34 responses</th>
<th>93 responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Important</td>
<td>30%</td>
<td>44%</td>
<td>56%</td>
</tr>
<tr>
<td>Important</td>
<td>54%</td>
<td>32%</td>
<td>34%</td>
</tr>
<tr>
<td>Neither</td>
<td>15%</td>
<td>15%</td>
<td>8%</td>
</tr>
<tr>
<td>Unimportant</td>
<td>2%</td>
<td>9%</td>
<td>2%</td>
</tr>
<tr>
<td>Very unimportant</td>
<td>0%</td>
<td>3%</td>
<td>0%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Learning Benefit 5 – They can allow learners to learn through experience in context</th>
<th>61 responses</th>
<th>35 responses</th>
<th>93 responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Important</td>
<td>43%</td>
<td>54%</td>
<td>71%</td>
</tr>
<tr>
<td>Important</td>
<td>54%</td>
<td>29%</td>
<td>27%</td>
</tr>
<tr>
<td>Neither</td>
<td>2%</td>
<td>11%</td>
<td>2%</td>
</tr>
<tr>
<td>Unimportant</td>
<td>2%</td>
<td>3%</td>
<td>0%</td>
</tr>
<tr>
<td>Very unimportant</td>
<td>0%</td>
<td>3%</td>
<td>0%</td>
</tr>
</tbody>
</table>
Discussion

These results provide some insight into why virtual worlds have not become a mainstream teaching tool and why some educators that have made use of them feel unable or unwilling to continue using them because of a range of issues across diverse areas including technical support, student attitudes and time pressures. Though virtual worlds were initially welcomed by educators amid great fanfare, this enthusiasm, reflected in the quick rise to the “peak of inflated expectations,” has not been maintained. New technologies can sometimes disappear forever after reaching this peak on the Gartner Hype Cycle (Lowendahl, 2013) and plunging deep into the “trough of disillusionment.” Perhaps the question to be asked in relation to 3D virtual worlds is whether the bottom of the trough has been reached, whether it is at the lowest point or is indeed on the way up to the “slope of enlightenment?” A recent development in education may give some insight: Tapson (2013) discusses the phenomenon of MOOCs (Massive Open Online Courses) in relation to the hype cycle and concluded that the MOOC phenomenon has climbed to the “peak of inflated expectations” very rapidly and will experience a short-lived “trough of disillusionment” before climbing the “slope of enlightenment” in the period between 2017 and 2022, then proceeding tsunami-like to overtake traditional university teaching once reaching the “plateau of productivity.”

With articles such as this one and those cited throughout, looking into the reasons people may not be using virtual worlds, we could soon see this information used to formulate best practice solutions. Most of the respondents discussed here are either using virtual worlds and intend to continue doing so or have never used them but may use them in the future. The data in this article also suggests that educators are not basing their intention to use virtual worlds on hearsay, hype or opinion, but on an understanding that there is sound pedagogy and educational merit in their use. For most, the reason they may not be using virtual worlds was focused around institutional concerns. Whether currently using virtual worlds or not, the respondents agreed that an institution that does not provide funding, technical or teaching support is the greatest barrier to the continued adoption of virtual worlds for teaching, learning and research.

Conclusions and looking to the future

Higher education institutions invest large amounts of funds in training staff to use new teaching platforms such as Learning Management Systems, for example, Moodle or Blackboard. Learning advisors skilled in the use of these technologies are available to assist those wishing to engage with the technology or to further resource their learning. This has not been the case with the adoption and use of virtual worlds. When new buildings or laboratories are built in the “real” world, they are seen as a concrete asset that should be used by successive groups of students, irrespective of which teacher is utilizing them. We suggest that as the assets in the virtual world are not obvious and not subject to space-utilization audits, they can sit vacant or are disposed of when a project finishes, rather than being re-used or re-purposed. It is imperative to have a sustainability plan in place if virtual worlds are to be a viable resource in future education. If one person is the instigator of the initiative and leaves the institution, there needs to be someone in place to take over the virtual world subscriptions and everything that comes with that academic’s work over the years. In order to keep the space and the classes that someone has established, there needs to be more than one person involved. So that each new project does not have to forge the exact same path, there needs to be a community of practice at each institution, even if that only consists of a small but accessible team. This team would preferably involve members that have the experience and ability to access documentation, resources and procedures needed to inform prospective users. With this type of resource in place and a policy to manage funding of virtual spaces, the authors agree that educators are ready to lift virtual worlds out of the Gartner trough and up the slope. Following the Gartner Hype Cycle, this would be an opportune time for the introduction of the third wave of virtual world solutions to help mitigate the barriers while ascending the slope.

Initial funding is often provided for the establishment and uptake of a virtual world for teaching and learning. However, provision is not made for ongoing technical support. This should be provided to update, back up, trouble-shoot and establish a “go to” person or group to curate the virtual space. Students need to be trained in how to use the virtual world for their learning. Academics should be trained in how to teach in a virtual world. This problem can be exacerbated by the increased casualization of the academic workforce and the delay in appointing casual teaching staff (just in time employment) so that it is not possible to educate sessional staff in how to use a 3D virtual world for teaching and learning in the time available. Often, there is no provision for the time it takes to upgrade a virtual world site or the expertise in the academic staff to undertake this, without additional training.

The authors contend that there is a future for teaching and learning in virtual worlds. The evidence outlined supports the notion that those who are teaching in a virtual world perceive these spaces are important for teaching and learning. Individuals and groups worldwide are undertaking research to provide empirical evidence of the value of teaching in a virtual world. Maybe in the light of this emerging evidence and with the issues mitigated by careful planning, resourcing and practical support solutions, the great educational potential of virtual worlds could be realized with them becoming a powerful tool in the arsenal of educators.
References


Coming Down to Earth: Helping Teachers Use 3D Virtual Worlds in Across-Spaces Learning Situations

Juan A. Muñoz-Cristóbal¹, Luis P. Prieto², Juan I. Asensio-Pérez¹, Alejandra Martínez-Monés³, Iván M. Jorrín-Abellán³ and Yannis Dimitriadis¹

¹School of Telecommunications Engineering, Universidad de Valladolid, Paseo de Belén 15, 47011 Valladolid, Spain // ²CHILI Lab, EPFL, Rolex Learning Center, Station 20, CH-1015 Lausanne, Switzerland // ³School of Computer Engineering, Universidad de Valladolid, Paseo de Belén 15, 47011 Valladolid, Spain // ⁴Department of Secondary and Middle Grades Education, Bagwell College of Education, Kennesaw State University, 1000 Chastain Road MD 0122. Kennesaw Hall, Bldg. #1, 30144-5591 Kennesaw, GA, USA // juanmunoz@gsic.uva.es // luis.prieto@epfl.ch // juuaase@tel.uva.es // amartine@infor.uva.es // ijorrina@kennesaw.edu // yannis@tel.uva.es

Corresponding author

ABSTRACT

Different approaches have explored how to provide seamless learning across multiple ICT-enabled physical and virtual spaces, including three-dimensional virtual worlds (3DVW). However, these approaches present limitations that may reduce their acceptance in authentic educational practice: The difficulties of authoring and sharing teacher-created designs across different 3DVW platforms, or the lack of integration of 3DVWs with existing technologies in the classroom ecosystem (e.g., widespread web-based learning platforms such as Moodle, or mobile augmented reality applications). Focusing on a specific kind of 3DVW (virtual globes, such as Google Earth, used like 3DVWs), we propose a system that enables teachers to deploy across-sites learning situations, which can be authored with a plethora of existing learning design tools, that involve different common web-based learning platforms, mobile AR applications and multiple kinds of virtual globes. A prototype of the architecture has been developed to evaluate this novel approach. The mixed-methods evaluation performed comprised both a feature analysis and a study where a teacher deployed an authentic across-sites learning situation including Google Earth used as a 3DVW. Such evaluation shows that the system enables teachers deploy learning situations over different technological ecosystems composed by physical and web spaces, as well as by 3DVWs.

Keywords

3D virtual worlds, Virtual globes, Google Earth, Augmented reality, Across-spaces learning

Introduction

Advances in Information and Communication Technologies (ICT) are bringing about new possibilities for learning, such as those involving different virtual and physical spaces. For example, activities related to botany using a web platform in the classroom can be complemented with activities in a nearby forest (Kurti, Spikol, & Milrad, 2008). Different approaches have explored how to provide a continuous learning experience in these across-spaces learning situations (Kurti et al., 2008; Muñoz-Cristóbal et al., 2014), thus moving toward “seamless learning” (Chan et al., 2006). Mobile devices and Augmented Reality (AR) are among the technical scaffolds that have been explored to connect these physical and virtual spaces (Billinghurst & Duensers, 2012; Muñoz-Cristóbal et al., 2014; Sharples, Sanchez, Milrad, & Vavoula, 2009).

Three-dimensional virtual worlds (3DVW) such as Second Life (http://secondlife.com) or Open Wonderland (http://openwonderland.org) constitute an additional type of learning space that can be found in currently proposed across-spaces learning scenarios. 3DVWs are three-dimensional virtual environments with similarities to the real world that provide the illusion of being there. 3DVW users are represented using avatars that can interact with other users and objects of the 3DVW in a synchronous or asynchronous fashion (Dickey, 2003; Warburton, 2009). The use of 3DVWs in education has been explored during the last decades, and has shown to provide different learning benefits. More specifically, 3DVWs increase student motivation and also enable the perception of objects from multiple perspectives, the simulation of experiences impossible in the real world, or help knowledge transfer to the real world through the contextualization of learning (Dalgarno & Lee, 2010; Dede, 2009; Dede, Salzman, & Loftin, 1996; Dickey, 2003; Warburton, 2009). Existing examples of across-spaces learning situations involving 3DVW include the combination of activities in Moodle (https://moodle.org) and Second Life (Livingstone & Kemp, 2008), or the synchronous interaction among students visiting a replica of a city in Open Wonderland and students physically located in the “real” city, using AR in mobile devices (Ibáñez, Maroto, García Rueda, Leony, & Delgado Kloos, 2012).
However, most of the approaches considering across-spaces learning situations that include 3DVWs show limitations that may contribute to the current lack of acceptance of their proposals in real educational practice (Gregory et al., 2013; Hendou, Limayem, & Thompson, 2008; Warburton, 2009). One limitation is the lack of support for teachers to create their own across-spaces learning situations. Also, the available range of technologies for the enactment of the authored scenarios is very limited (e.g., the specific combination of Moodle and Second Life in Livingstone & Kemp, 2008, or Open Wonderland and an ad-hoc mobile client in Ibáñez, et al., 2012). Additionally, existing proposals tend to consider 3DVW-supported activities in a rather isolated way with respect to activities in other 3DVWs, or supported by already existing technologies in the classroom (VLEs, Web 2.0 tools, AR tools).

To overcome these limitations, this paper proposes the architecture and prototype of a system capable of supporting teachers in creating, with a number of existing authoring tools, and deploying their own across-spaces learning situations in a variety of technological ecosystems comprising multiple learning spaces. These ecosystems may be composed of different mainstream VLEs and Web 2.0 tools (web learning space), multiple mobile AR applications (augmented physical learning space), as well as distinct 3DVWs (3DVW learning space). Thus, the system enables activities taking place in multiple physical, 3DVW and web spaces, at the same time or sequentially. Additionally, learning designs can be shared and reused in different technological ecosystems (e.g., those including different 3DVWs).

The system proposed is an extension of the one reported in (Muñoz-Cristóbal et al., 2014), which did not support 3DVWs as learning spaces. Our new proposal integrates Virtual Globes (VGs) such as Google Earth (http://www.google.com/earth/), used as a 3DVW. VGs are virtual 3D representations of the surface of the Earth, which are widely known and recurrently used with educational purposes (Chen & Choi, 2010; Lund & Macklin, 2007; Rakshit & Ogneva-Himmelberger, 2008; Schultz, Kerski, & Patterson, 2008; Ternier, Klemke, Kalz, van Ulzen, & Specht, 2012; Wells, Frischer, Ross, & Keller, 2009). Although VGs lack certain presence and interaction features of 3DVW, there exist proposals for their conversion into 3DVWs (see, e.g., Dordevic & Wild, 2012, or http://youbeq.com, which include user interaction and avatars in Google Earth). Technical reasons also make it recommendable to integrate VGs in across-spaces learning platforms, since both VGs and physical spaces use the same type of (geographical) coordinates, thus simplifying the flow of learning artifacts and participants among spaces. Furthermore, the growing availability of 3D content for mainstream VGs (see, e.g., Xiao & Furukawa, 2012) opens new opportunities for setting up innovative learning scenarios in VGs, and may promote the adoption of VGs used like 3DVWs by teachers.

The next section describes limitations of current approaches to across-spaces learning situations that include 3DVWs, and distills design requirements that a system should fulfill in order to enable teachers to devise and perform such situations. Then, the paper proposes the architecture and prototype of a system implementing those design requirements. The paper also reports on the evaluation of the proposed system, which involved a feature analysis and a study where a teacher designed and deployed an across-spaces learning situation. Finally, some reflections and conclusions are mentioned.

**Limitations of current approaches and design requirements**

Several proposals for supporting across-spaces learning situations include activities in 3DVWs. In order to connect *3DVW with web spaces*, some approaches (Dickey, 2003) simply embed a web browser in the 3DVW user interface, triggering different web pages upon the occurrence of certain events in the virtual world. Some approaches can display web pages inside the 3DVW. This is also the case of OPENET4VE (Fernández-Gallego, Lama, Vidal, Sánchez, & Bugarín, 2010) which is able to deploy learning situations in different 3DVWs. Other interesting proposals are Sloodle (http://www.sloodle.org; Livingstone & Kemp, 2008), for linking Moodle with Second Life and OpenSim, and the system presented by Pourmirza & Gardner (2013), which links Facebook (https://www.facebook.com) with Open Wonderland.

Other approaches study the connection of *3DVWs with physical spaces*. In the Citywide project (Izadi et al., 2002), students explore a physical space (e.g., an archaeological site). At specific locations, students are asked to connect to a 3DVW where they can access objects related to the physical locations. Similarly, Ibáñez, et al. (2012) present a hybrid learning environment where activities may occur both in a physical street using mobile AR, and in a 3DVW...
mirroring that same street using desktop computers. Participants in any of the two spaces can see avatars of the other participants and interact with them.

Table 1. Limitations of existing approaches for adoption in educational practice, and design requirements proposed to overcome such limitations

<table>
<thead>
<tr>
<th>Limitations</th>
<th>Design requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Limited range of supported spaces (physical, web, 3DVW).</td>
<td>DR1. Virtual learning resources should be accessible from all three kinds of spaces: physical, web, 3DVW.</td>
</tr>
<tr>
<td>Lack of authoring tools usable by teachers.</td>
<td>DR2. Allow teachers to create across-spaces learning situations without requiring high level of technical knowledge.</td>
</tr>
<tr>
<td>Activities designed for a specific 3DVW cannot be reused in other virtual worlds.</td>
<td>DR3. Enable teacher designs to be deployed in multiple 3DVWs.</td>
</tr>
<tr>
<td>Isolation of 3DVWs from other virtual worlds and technologies in use by teachers.</td>
<td>DR4. Compatibility with multiple platforms/systems within each supported space (physical, web, 3DVW).</td>
</tr>
<tr>
<td></td>
<td>DR5. Users in a 3DVW and users accessing other 3DVWs or platforms/systems on other spaces, have to be aware of the presence of each other.</td>
</tr>
<tr>
<td></td>
<td>DR6. Allow the use of different enactment technologies already in use by teachers and institutions.</td>
</tr>
</tbody>
</table>

Other proposals explore across-spaces learning situations involving *three types of spaces: physical, web and 3DVWs*. DigitalEE (Okada, Tarumi, Yoshimura, & Moriya, 2001) is a system for collaborative environmental education wherein a 3DVW replicates a forest. The participants in the 3DVW can interact with others moving through the real forest, equipped with laptops and GPS. Participants in the physical forest can generate and interact with information (e.g., pictures, videos) which is uploaded to the 3DVW as HTML pages.

However, these proposals are affected by some generic limitations of the 3DVWs (Gregory et al., 2013; Hendaoui et al., 2008; Warburton, 2009), which may prevent the adoption of such approaches in educational practice. In the following paragraphs, we indentify limitations that can affect especially across-spaces learning situations involving 3DVWs. From these limitations, we derive a list of design requirements (DR) that may help to overcome such limitations (Table 1).

Typically, across-spaces approaches limit their range of applicability to a *specific combination of physical, 3DVW or web spaces* (see *DR1* in Table 1). Of the aforementioned proposals, only DigitalEE supports learning activities in a 3DVW as well as in physical and web spaces. Thus, e.g., if a teacher using Sloodle wants to include an activity in a physical space, she would need to use another system, probably not integrated so seamlessly with the rest of the technological support.

In addition, there are few cases in which the design of learning scenarios is supported by *authoring tools usable by teachers* without a high level of technical expertise (see *DR2* in Table 1). Such authoring tools have been studied in the field of learning design (Koper, 2005), as a way of explicitly represent pedagogical ideas using computer-interpretable languages, sometimes independently from the targeted enactment platform. OPENET4VE and the system described by Ibáñez, et al. (2012), for example, enable the use of IMS-LD (IMS Global Learning Consortium, 2003) learning designs, thus allowing the use of different authoring tools based in that specification to create activities. In a different way, teachers may design learning situations with Sloodle by using Moodle’s user interface.

Another common limitation is that *activities designed for a specific 3DVW cannot be reused in other virtual worlds* (see *DR3* in Table 1). Of the aforementioned approaches, only OPENET4VE and Sloodle enable the use of more than one 3DVW (Second Life and OpenSim), and DigitalEE provides limited support to other VRML-compliant 3DVWs.

Finally, in existing proposals, *virtual worlds normally are isolated, disconnected from other 3DVWs and technologies already in use by teachers and institutions* (e.g., VLEs, Web 2.0 tools or AR applications), thus
hampering the adoption of such proposals in existing educational practice (see DR4, DR5 and DR6 in Table 1). Among the aforementioned approaches, only Sloodle integrates a widespread VLE (Moodle), and the system presented by Pourmizra & Gardner (2013) integrates Facebook. The rest of approaches integrate mostly ad-hoc, not widely adopted platforms.

Proposal

Following these design requirements, we propose an architecture that enables teachers to define their across-spaces learning situations with different existing authoring tools, and deploy them in complex technological settings which include multiple VLEs, mobile AR clients and VGs used like 3DVWs.

Architecture

Previous work by the authors (Muñoz-Cristóbal et al., 2014) has tackled similar problems of connecting isolated spaces using existing platforms, enabling teachers to deploy their learning situations in different web and AR-enabled physical environments, but not in 3DVWs. Following the design requirements in the previous section, we now extend the previous proposal to consider also VGs used like 3DVWs. The architecture (Figure 1) is based on adapters which allow the integration of multiple elements of the same type (i.e., authoring tools, external web tools, VLEs, mobile AR clients and VGs). Once an element is integrated in the architecture, it is connected with the rest of the elements, enabling a multi-to-multi approach comprising already-existing authoring tools (compliant with DR2) (e.g., those described in The Learning Design Grid, 2013), VLEs (e.g., Moodle or Blackboard [http://www.blackboard.com]), mobile AR clients (e.g., Junaio [http://www.junai0.com] or Layar [https://www.layar.com]) and VGs (e.g., Google Earth or SkylineGlobe [http://www.skylinesoft.com]) (DR3, DR4, DR6). Thus, a teacher could use multiple learning design authoring tools to define her across-spaces activities, and complete them (e.g., with positioning information for the learning artifacts in physical spaces and VGs) in the graphical user interface (GUI) of the system (DR2).

The architecture integrates also multiple tools and resources by means of Tool adapters. These Tool adapters may integrate, for instance, different Distributed Learning Environment (DLE) (MacNeill & Kraan, 2010) proposals such as Glue! (Alario-Hoyos et al., 2013) or IMS-LTI (IMS Global Learning Consortium, 2012). Thus, a single DLE adapter allows the use of multiple tools and resources in a learning scenario (e.g., in the case of Glue!, different Web
2.0 tools, widgets, 3D models or web resources) (DR4, DR6). These tools and resources can be accessed from any of the VLEs, mobile AR clients and VGs integrated in the system (DR1).

Since VGs are not by themselves 3DVWs (they do not provide avatars or interaction between users), the architecture, by means of the adapters and the Manager, enables the use of VGs as if they were 3DVWs, by adding avatars for the users and allowing interaction with any virtual resources and tools supported by the architecture (DR1, DR5). Students can interact with resources and tools or create new ones (e.g., opening a Google Docs document located in a VG, or uploading and positioning in the VG a picture taken in the physical space with their mobile device). Also, different communication tools (e.g., a chat) can be included in the teacher’s design and be deployed to this VG-based 3DVW, thus enabling students in a VG to chat with students using mobile AR in a physical location, or with students using a VLE in the classroom. The Manager element acts as a central hub for synchronizing user information in all spaces. Thus, e.g., users in a physical location can see the avatars of users in a VG using AR, and vice-versa (DR5).

Integration of multiple Virtual Globes

The integration of an existing VG in the architecture depends on a single requirement: the target VG should have an API or SDK for allowing third parties to position data elements (e.g., learning artifacts or avatars), typically known as “points of interests” (POI). Our proposal uses a POI model for learning artifacts and avatars, which is based on the model described in (Muñoz-Cristóbal et al., 2014). VG adapters are in charge of transforming POI representations following this common model, to the different native representations used by each of the VGs.

In order to assess the feasibility of such transformation for existing VGs, we have studied the API or SDK of some of the most widespread VGs (Rakshit & Ogneva-Himmelberger, 2008; Schultz et al., 2008): Nasa World Wind (http://worldwind.arc.nasa.gov), Bing Maps (http://www.bing.com/maps/), SkylineGlobe, Google Earth, Google Maps/Street View (https://maps.google.com) and ArcGIS 3D Analyst (http://www.esri.com/software/arcgis/extensions/3danalyst). We consider Google Maps/Street View also a VG because, although it does not support 3D views, its 360º realistic pictures provide an experience very close to that of 3D interfaces. We have also included Bing Maps for the sake of completeness, although its current support for 3D views is quite limited. Figure 2 shows the POI model used by the proposed system, and the elements of such model that are supported in each VG. The figure also shows how transforming POIs between the proposed system and the different VGs would not imply a significant loss of information. Therefore, all VGs reported in Figure 2 seem susceptible of being integrated into our proposal (DR3, DR6).

Figure 2. Elements of the POI model employed by the proposed system and their correspondence with the POI model employed by different VGs. (Position-type attribute is not supported, since VGs use only geographical coordinates)
Figure 3. Example of a learning situation illustrating the user interfaces of different learning environments: Google Earth (a) and Google Maps/Street View (b) used like 3DVWs, the Junaio AR client (c) and Moodle (d).

Prototype

We have developed a prototype of the architecture, extending the predecessor architecture’s prototype (Muñoz-Cristóbal et al., 2014), which already integrated three existing authoring tools (as well as other authoring tools compliant with IMS LD level A), two VLEs, four mobile AR clients and the Glue! DLE adapter. Furthermore, we have now integrated two VGs: Google Earth and Google Maps/Street View, used as 3DVWs, modifying the Manager to synchronize user information (and therefore, their avatars) in mobile AR clients and VGs. Figure 3 illustrates the user interfaces of different learning environments in an example learning situation involving Google Earth (a) and Google Maps/Street View (b) used like 3DVWs, the Junaio AR client (c) and Moodle (d). In the figure, a student (student_a11) may access the same activity in virtual downtown Valladolid (Spain) from two VGs: Google Earth (a) and Google Maps/Street View (b). From both VGs, she can see the avatar of another student (student_a12) and some geopositioned tools (in this case, a chat and a Google Docs document). Student_a12 is physically in the street in downtown Valladolid, and he is also watching the avatar of student_a11 and the geopositioned tools (c), using a mobile device with Junaio AR client. The chat shared by these students is also accessible from within Moodle (d).
Evaluation

An evaluation has been conducted to explore the research question driving our work:

*Does the proposed system enable teachers to deploy their across-spaces learning situations including web spaces, AR-enabled physical spaces, and VGs (used like 3DVWs)?*

The evaluation consisted of (1) a study wherein a university teacher used the system to deploy an across-spaces learning situation that included Google Earth used like a 3DVW; and (2) a feature analysis, where the evaluation team scored the support of existing approaches and the proposed system to the design requirements defined in Table 1. The next subsections describe the methodological considerations, the evaluation process and its results.

Method

We have followed the Computer Supported Collaborative Learning - Evaluand-oriented Responsive Evaluation Model (CSCL-EREM) (Jorrín-Abellán, Stake, & Martínez-Monés, 2009), using a variety of data gathering techniques in a mixed-method approach (Creswell, Plano Clark, Gutmann, & Hanson, 2003). CSCL-EREM is a framework, focused on the phenomena under evaluation, that provides evaluators with concepts and tools to guide the evaluation of CSCL phenomena (in this case, a technological innovation) in ubiquitous collaborative learning settings. CSCL-EREM is based on the responsive evaluation approach (Stake, 2004) and, therefore, aims at responding to the participants in the evaluation (instead of just describing, measuring or judging them), to get a deep understanding of the setting that may facilitate the adoption of the innovation in practice. This evaluation method is framed within the interpretive research perspective (Orlikowski & Baroudi, 1991), that does not pursue statistically significant results or generalizations, rather aiming a deeper understanding of the concrete phenomena under study (Guba, 1981), in our case, the use of the proposed system by a teacher.

![Figure 4. Anticipatory data reduction analysis. (RQ) Research question. (I) Issue. (IQ) Informative question.](image)

To help illuminate our research question, we performed an anticipatory data reduction process (Miles & Huberman, 1994) during the evaluation design (see Figure 4). We defined an issue as the main conceptual organizer of the evaluation process. This issue was split into two more concrete topics, to help us understand different relevant dimensions within the issue: *The deployment affordability for the teacher of her across-spaces learning situations*
As mentioned, the evaluation consisted of a study and a feature analysis. Along the study, a profuse set of data gathering techniques and data sources has been used: teacher-generated artifacts (e.g., emails or learning designs), time recordings, screen recordings (with software that recorded operations in the screen as well as audio and video out of the screen), naturalistic observations (audio, video, pictures and observation notes), web-based questionnaires and interviews (see Figure 5). In parallel with the study, we performed a feature analysis: a systematic comparison of the proposed architecture with alternative approaches, in order to explore if the support of the proposed system to the design requirements defined improves the existing approaches’ support. For the feature analysis, we followed the screening method of the DESMET evaluation methodology (Kitchenham, Linkman, & Law, 1997). The screening method is a qualitative feature-based evaluation performed by a single individual or an evaluation group, who not only determines the features to be assessed and their rating scale but also does the assessment. Questionnaires (Score Sheets in Figure 5) are used to assess the features, and the scores are compiled in a final report (Evaluation Profile).

During our evaluation, triangulation of methods, data sources and evaluators was used, to cross-check data and interpretations as well as to assure the quality and credibility of the research (Guba, 1981). Figure 5 shows the evaluation process, which is divided in different happenings (evaluation events), as the CSCL-EREM model recommends, as well as the different data gathering techniques used in each happening, indicating the labels used to refer to them throughout the text.

![Figure 5. Evaluation happenings and data gathering techniques](image)

**Study**

**Context**

The study was performed in the College of Education and Social Work, University of Valladolid (Spain). In the first year (out of four) in the Degree in Primary Education, within a mandatory course on ICT for pre-service teachers, four university teachers, usually perform an across-spaces learning situation related to the learning effects of advertising in everyday life. Usually, the situation involves the wiki of the course (used as a VLE), activities in the streets using mobile devices, and activities using Google Maps. Students are instructed to capture pictures of advertisement panels in their way home, writing down their location. Later on, they have to upload the pictures
manually to the wiki and, in groups, they create a map in Google Maps, marking the routes followed by the group’s members, and create a marker in the map for each picture, associated to the URL of the picture in the wiki. Finally, students have to elaborate a reflective critique about the advertisements.

**Intervention**

One of the course teachers, a pedagogue relatively new to teaching (four years of teaching experience), who is familiar with ICT tools, showed interest in enhancing their usual across-spaces learning situation using Google Earth as a 3DVW (instead of the simple 2D Google Map described above). The proposed system may also help improve the connection of activities across the different spaces, automating several manual operations. In a first happening (H1, Figure 5), the teacher conceptualized a learning design without the support of an authoring tool. Then (H2), an ICT-expert from the evaluation team used the WebCollage (Villasclaras-Fernández, Hernández-Leo, Asensio-Pérez, & Dimitriadis, 2013) authoring tool to represent the teacher’s learning design, and used the system prototype to deploy the design across the technological enactment platforms (wiki, QR codes and Google Earth). In a subsequent happening (H3), the participant teacher (with the support of an ICT-expert evaluator) performed the authoring of the design with WebCollage and the actual deployment using the system prototype. The teacher also assessed the resulting technological infrastructure, using it in the role of a student. During this deployment session, three observers were present (taking notes, pictures, videos and audio) who, later on, created a multimedia collaborative triangulated observation report, including also an annotated analysis of the complete video of the session. Finally (H4), the teacher gave feedback by means of a web-based questionnaire and a semi-structured interview (to provide further details about the deployment process and her perception of it).

**Findings**

Figure 6 illustrates the learning situation designed by the teacher, as a variation of the original across-spaces activities usually performed. It is composed of four collaborative activities spanning four hours of face-to-face work and two hours of remote work, performed in the classroom (using a wiki and Google Earth), in the streets (using QR codes and mobile devices for reading them), and online (using Google Earth). In order to use some of the time saved (due to the system’s automation of the enactment), a new “counter-ad” activity was included at the end of the design. Also, a simple analysis of advertisement in general was replaced by a detailed analysis of the student-contributed advertisements.

![Figure 6](image.png)

*Figure 6. Description of the learning situation created by the teacher*

The teacher deployed her learning situation using the prototype (IQ 1.1, see Figure 4), taking 54 minutes to author the design in WebCollage, and 37 minutes to configure that design using the prototype user interface (positioning learning artifacts, reusing tools between activities, etc.) [Time 3, Screen 2, Obs 1] (see Figure 5). As expected, the evaluator spent considerably less time than the teacher in this process, especially regarding WebCollage (7 minutes in WebCollage and 12 in the prototype [Time 2, Screen 1]). This difference was not only due to the evaluator’s higher expertise with the systems, but also due to the mechanical nature of copying already-designed activities rather
than reflecting on them (as the teacher did) [Screen 1]. Indeed, even if the conceptual design had already been done on paper, the teacher included last-minute changes in activities during her use of WebCollage (e.g., a questionnaire was replaced by the counter-ad activity in Google Earth [Teacher 1, Screen 2]), and she queried the ICT-expert evaluator about the modeling of some activities using the provided tools ("... The teacher explains a new activity for the last part of the design where she wants students to generate a counter-ad based on one of the previously analyzed ones. ‘How can I make it happen? Which is the best way for my students to upload the image with the counter-ad to Google earth?’ [...] [Obs 1]). The teacher considered the total time dedicated to the process was “long” (about 90 minutes) (IQ 1.2) ([when asked to assess the time dedicated in the deployment session:] “Well, I think it was a long time [...] because in the end it took an hour and a half” [Int 1]). Nevertheless, she considered such time as acceptable since she felt that the system was intuitive, and subsequent deployments would take less time (“I think that with another go of practice, it would take me a lot less time” [Int 1]; “I think [the time dedicated] is acceptable, since the design can be reused in upcoming years, and in other courses performing minor adaptations [...] anyway, I’m sure that, with practice, the time [it takes] would be greatly reduced, since it is quite intuitive once you know some of the terms that are confusing at the beginning” [Quest 1]). The teacher also considered the deployment process affordable (IQ 1.3), both in a questionnaire ([to the assertion “I think that the deployment of the scenario with the system has been easy”] she answered “Agree”, 5 in a 1-6 scale [Quest 1]) and in the interview [Int 1]. She also showed interest in using the system in her real practice (IQ 1.5). This idea was confirmed during the deployment session (“I want to use it” [Screen 1]) as well as in a questionnaire (“I have the idea of putting [the scenario] in practice the next semester in my courses” [Quest 1]) and on the interview (“ [...] and probably, if we put this [learning situation] into practice, since I’m convinced that I want to put it into practice, we will change things again [...]” [Int 1]). After the deployment, she reviewed the resulting learning environment in the wiki and in Google Earth (see Figure 7), and simulated the tasks of a student [Obs 1, Screen 2]. She confirmed that it corresponded well to her initial design (IQ 1.6) [Int 1, Quest 1] (e.g., [to the assertion “do you consider that the result of the deployment process reflects your designed learning situation?”] she answered “Strongly agree”, 6 in a 1-6 scale). Finally, she valued as important the three features that she had used (IQ 2.7, IQ 2.8, IQ 2.9), both in the questionnaire (she answered “Strongly agree”, maximum in the 1-6 scale, to the three questions asking about those features’ importance [Quest 1]), and in the interview (e.g., “[...] a teacher has to be trained to be able to design contextualized activities [...]”, and that implies going beyond things or recipes that already exist. She has to be able to design these materials with the things that she can reach. On the other hand, a teacher also has to reuse things that are known to work, and which are free, [Web] 2.0 tools available to everybody” [Int 1]).
Feature analysis

Simultaneously to the study, a feature analysis was performed by the evaluation team (H5, see Figure 5). Table 2 shows the Evaluation Profile summarizing the scores obtained (in a 0-5 scale) by the different existing approaches reported in the literature, assessing their support to the design requirements mentioned in Table 1. The publications on each existing approach were studied by the evaluation team. Then, evaluators provided an individual score [Score 1], which was then shared in a two hours panel to discuss conflicting criteria and agree on a final score for each of the approaches.

As Table 2 shows, the proposed system supports all the defined design requirements (IQ 2.1 – IQ 2.6). The feature with the lowest degree of support by the proposed system is DR2, since the authoring tool and user interfaces of the different systems (both the prototype and the authoring tools it supports) could be improved to reduce the learning curve for teachers without high technical knowledge. The existing proposal that was scored nearest second was DigitalEE, since it supports all three spaces (web, physical and 3DVW), and has potential to integrate more than one 3DVW (since it used a standard for the 3D scene’s models).

Table 2. Evaluation profile of the different approaches: (Dickey, 2003) (Di), (Fernández-Gallego et al., 2010) (OP), (Livingstone & Kemp, 2008) (Sl), (Pourmirza & Gardner, 2013) (PG), (Izadi et al., 2002) (Cw), (Ibáñez et al., 2012) (Ib), (Okada et al., 2001) (DEE) and the proposed system (PS)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Conformance score obtained</th>
</tr>
</thead>
<tbody>
<tr>
<td>DR1. Accessibility of learning resources from all three kinds of spaces.</td>
<td>Di 0 OP 0 Sl 0 PG 0 Ib 5 DEE 5 PS 5</td>
</tr>
<tr>
<td>DR2. Allow teachers to create across-spaces learning situations.</td>
<td>1 3 3 3 3 3 3 3</td>
</tr>
<tr>
<td>DR3. Enable teacher designs to be deployed in multiple 3DVWs.</td>
<td>0 4 3 0 0 3 4 4</td>
</tr>
<tr>
<td>DR4. Compatibility with multiple platforms/systems within each supported space.</td>
<td>0 3 0 0 0 0 0 5</td>
</tr>
<tr>
<td>DR5. Presence awareness between users in a 3DVW and users accessing other 3DVWs or platforms/systems on other spaces.</td>
<td>0 0 4 0 4 4 4 4</td>
</tr>
<tr>
<td>DR6. Allow the use of different enactment technologies already in use by teachers and institutions.</td>
<td>1 3 3 3 0 1 1 4</td>
</tr>
<tr>
<td>Total</td>
<td>2 13 13 6 7 8 16 25</td>
</tr>
<tr>
<td>% over the total possible</td>
<td>7 43 43 20 23 27 53 83</td>
</tr>
</tbody>
</table>

Discussion, conclusions and future work

The data gathered in our evaluation highlight that the deployment using the system is affordable for a teacher (topic 1). In addition, the feature analysis shows how the proposed system supports the design requirements defined to ease the adoption of across-spaces learning situations including 3DVWs. To the best of our knowledge, the rest of systems analyzed do not support all such design requirements. Indeed, none of the existing approaches supports all three features that were actually used by the teacher in the design and deployment of her learning situation (topic 2). Thus, the learning scenario devised by the teacher would not be deployable by her without the support of the proposed system (or at least, not in 90 minutes). The findings across these two topics help illuminate our initial research question, providing evidence that the proposed system enables teachers to deploy their across-spaces learning situations including web spaces, AR-enabled physical spaces, and VGs used like 3DVWs. We expect that the system and the design requirements proposed here may help to promote the adoption of 3DVWs in across-spaces learning situations, moving 3DVWs closer to everyday practice of teachers and their institutions. We claim that using currently-widespread VGs (e.g., Google Earth) as 3DVWs, and achieving seamless integration with other existing educational technologies (authoring tools, VLEs, Web 2.0 tools, 3D models or AR applications), could significantly help adoption in real practice.

It is worth noting that our feature analysis does not attempt to compare or evaluate across-spaces approaches in general. Rather, it is only valid to compare how different approaches support the specific design requirements we have defined as interesting for teacher adoption. Additionally, the feature analysis was performed using the cited publications as its main base. The evaluation team could not actually test all the approaches or study more detailed documentation. The feature analysis could also been improved by using other methods described by Kitchenham, et al. (1997).
We intend to delve further into this line of research, enacting the learning situation designed by the teacher, as well as exploring other learning situations that make use of the aforementioned set of features, with different teachers, in different educational contexts, and using a variety of existing technologies (e.g., other VLEs, mobile AR clients and VGs). We also plan further research to improve the prototype’s usability (DR2), and its current limitations regarding identity and immersion of the users in the 3DVW (Dalgarno & Lee, 2010) (e.g., improving avatar features). Further investigation is also needed to propose a more general architecture, that is able to integrate other types of 3DVWs (e.g., Second Life), or to explore other known limitations of the 3DVWs (e.g., technological, social or psychological) not studied in the present article.

Acknowledgements

This research has been partially funded by the Spanish Project TIN2011-28308-C03-02 and the European Project 531262-LLP-2012-ES-KA3-KA3MP, as well as by a Marie Curie Fellowship within the 7th European Community Framework Programme (MIOCTI, FP7-PEOPLE-2012-IEF project no. 327384). The authors thank especially the teacher involved in the evaluation, as well as Mario Marinero, who developed a first version of the Google Earth adapter in his Degree Thesis and Javier Hoyos, who developed the current version of the Google Earth and Google Maps/Street View adapters. The authors thank also the rest of the GSIC/EMIC research team for their ideas and support.

References


Toward Educational Virtual Worlds: Should Identity Federation Be a Concern?

Gonçalo Cruz1,2*, António Costa2, Paulo Martins1,2, Ramiro Gonçalves1,2 and João Barroso1,2

1INESC TEC (former INESC Porto), 4200-465 Porto, Portugal // 2School of Science and Technology, Universidade de Trás-os-Montes e Alto Douro, 5000-801 Vila Real, Portugal // goncaloc@utad.pt // acosta@utad.pt // pmartins@utad.pt // ramiro@utad.pt // jbarroso@utad.pt

*Corresponding author

ABSTRACT

3D Virtual Worlds are being used for education and training purposes in a cross-disciplinary way. However, its widespread adoption, particularly in formal learning contexts, is far from being a reality due to a broad range of technological challenges. In this reflection paper, our main goal is to argue why and how identity federation should be discussed and adopted as a solution to several barriers that educators and institutions face when using Virtual Worlds. By presenting a clear set of scenarios within different dimensions of the educational process, as classroom management, content reuse, learning analytics, accessibility, and research, we consider identity, traceability, privacy, accountability, and interoperability as main concerns in order to support our argument. Finally, we conclude the paper by presenting paths to a proposal for a workable solution, through the analysis and reflection of different and current efforts that has been made by other teams, towards future technological developments.

Keywords

Digital Identity, Identity Management, Federated Identity, Virtual Worlds, Education

Introduction

We live the most collaborative experience since the beginning of the World Wide Web, due to the easy manner in which one can connect with other users, resources, services and information. To attain things such as books we wish to buy, tuition fees we have to pay, or even the completion certificate of an online course we attended, we constantly face the need to use our identity, and ensure to other entities that they are indeed dealing with us. This raises ethical and legal questions about privacy, security and accountability, both for users and entities - in the use and sharing of their personal information.

Thus, it's strictly necessary to have management tools available that enable not only the certification and validation of users’ identities, but also assure them that there will be a basic level of privacy when they share their personal information with other parties. This is why the existence of federations between entities, through the mutual agreement and establishment of policies, practices and common standards, proves to be of greater importance. The inherent advantages of these trust relationships are numerous and constantly cited on literature (Madsen et al., 2005; Shim et al., 2005; Smith, 2008).

In education, identity federation has also been a topic of interest for researchers (Linden, 2005; Hammerle, 2006; Aguirre et al., 2008), but it mainly addresses the use of web-based systems such as wikis, learning and content management systems, forums, academic portals, repositories, etc. Regarding 3D Virtual Worlds, some studies were developed within this field with emphasis on other perspectives, such as sociology (Lorentz, 2011), anthropology (Gabriels et al., 2011), and psychology (Aas, 2011). However, other studies show identity issues as barriers and challenges for students, staff and institutions in the adoption of Virtual Worlds for learning (Palomaki, 2009; Dalgarno et al., 2011), but we are left with a lack of clear evidence that supports the argument on why identity federation should be taken into account as one of the major concerns towards the widespread use of Virtual Worlds in education.

For that purpose, there is a need to understand in which dimensions of the educational process, identity federation technology has the potential to bring benefits. Students, teachers and institutional leaders feel somewhat unmotivated, unconfident and insecure in the development, management, support and assessment of classrooms or educational projects in Virtual Worlds due a broad range of factors related to identity issues. From time consumption (e.g., proliferation of students’ and teachers’ logins, passwords, roles and other data for different institutional systems, including Virtual Worlds; track, link and assess huge amounts of 3D data, as students' in-world access, classroom attendance and task progress; etc.), to costs (e.g., 3D resources development, interoperability and
reutilization, copyright policies and licensing terms, etc.) and reputation (e.g., ethical and legal questions related with avatars’ problematic behaviors, as inappropriate gestures and customizations, in which privacy assurance and users' accountability are compromised).

In these and other dimensions that we will present afterwards, several identified barriers and challenges can be solved and/or minimized by identity federation technology. However, other ones will remain without a clear solution for teachers and institutions. For instance, in 3D distance learning classrooms, if on the one hand we can manage students’ attendance based on their avatars’ status, behavior and interaction during the class, logins’ and logouts’ time and zone, etc., on the other hand we can't fully ensure if the student behind the avatar is really who we think it is. That is a classical challenge, throughout the years, common to any social or group application to be implemented in the classroom, not just Virtual Worlds.

Although, besides the argument on why identity federation technology should be used towards the massive adoption of Virtual Worlds in education, we also need to discuss how the solution must be figured out. Therefore, we begin by describing the state of the art related to the digital identity topic, framing it within the e-learning and distance education fields. Subsequently, we move forward to the major contribution of our work, clearly demonstrating a set of different problematic scenarios that underlie our argument, according to the current solutions and efforts. Finally, we end the paper by proposing paths to a practicable solution, in order to guide future technological developments that can solve the different problems identified within the presented educational scenarios.

**E-Learning and digital identities**

Digital identities are electronic representations of known information about a specific individual or organization, such as attributes, traits or preferences (Anwar et al., 2006). In e-learning, related with the concept of telepresence and how an individual manifests electronically, digital identities are created not only by the inclusion of biographical details but also with information on how the individuals communicate and interact with others in a shared online space.

Distance education may be more conducive to academic dishonesty than face-to-face instruction (Baron & Crooks, 2005), and the large gap existing between the current generation of students and teachers may mean more opportunities for it to occur (Windham, 2005). Thus, much effort has been put into the effective verification of students’ identities through online learning authentication (Bailie & Jortberg, 2009). In order to ensure academic integrity (Roberts & Hai-Jew, 2009), and prevent negative behaviors such as cheating (Bedford et al., 2009) and plagiarism (Olt, 2009), institutions need to verify if the current person that signed up for the learning is the one taking the course. Also called accountability, this is of critical importance for the reputation of the institutions, and defined as an encouraging process for students to become more involved in managing the risks they face (Whitson, 2009).

There are different types of authentication methods, depending on the data requirements used for that purpose. Some systems use only a password and others use smart cards or biometrics (physiological characteristics and behaviors), in order to link an individual with his/her own digital identity. Common biometrics measures involve the uses of fingerprints, iris scans, facial scans and voice (Ahmed & Moskowitz, 2005), and others use a combination of information, such as keystroke latencies in word-processing (Joyce & Gupta, 1990), search interests (Rowe, 2010), or even traffic-based access network (Barisch, 2009). At a very basic level, many systems use Internet Protocol (IP) address tracking to link individual users with their respective locations; IPs and real users are closely related (Clauß & Schiffner, 2006). For instance, logging in with a particular identity from various IPs in geographically dispersed locations may be one way of identifying concerns.

Students have lots of personal information that they share in online courses or activities. They share personal information within their work and the access to that information should be protected. It may be wholly private, disseminated among a small group of learners, or published on the Web, but whatever the case, the student has intellectual property rights over his/her own work that should also be preserved. In this sense, the concepts of authoring and privacy are strictly related to the knowledge of identity and should be taken into account by educators and institutions. Sometimes, a pseudonym may be used as a unique identifier, in order to provide an anonymous
identity during specific learning activities (e.g., peer reviewing). The pseudonym is an identifier where only the entity that assigned this type of identification knows the real world identity behind it (Anwar and Greer, 2012).

In the specific case of 3D Virtual Worlds, some researchers have found a very disconnected relationship between the avatar's information and the user's real world identity (Junglas et al., 2007), which suggests a greater range of identity differentiation. Such variant identities may allow them to experiment a set of roles and opportunities for social interaction. Many authors are against this digital identity anonymity because of the fear that anonymity would result in misdeeds or other threats, especially in business and educational contexts (Adrian, 2008; Boon & Sinclair, 2009).

Identity and access management

The concept of access management is used to describe the process of allowing access to protected online information. It mainly describes the administrative procedure to allow access to online resources or services, for any individual based on the provided identity (Alves & Uhomoibhi, 2010). It includes authentication (the process of verifying the identity of any individual that is requesting access), authorization (the process of determining what kind of access that should be granted to the user based on his/her credentials), attributes (information about the user, as membership or role), and trust (the agreement between different parties and systems for sharing identity information).

There are several ways in order to control user access to resources. One of the most known is Role-Based Access Control (RBAC). A method that regulates access to computer or resources based on the roles or privileges of individual users within an institution (Onashoga et al., 2014). In this context, access is the ability of an individual user to perform a specific task, such as view, create, or modify a file. Roles can be defined according to the user skills, authority, and responsibility within the institution. When properly implemented, it enables users to carry out a wide range of authorized tasks by dynamically regulating their actions according to flexible functions, relationships, and constraints. Not only services based on the Web, which are frequently used in e-learning contexts (e.g., wikis, learning management systems, forums, etc.), but also 3D Virtual Worlds (e.g., Second Life and OpenSimulator) use this type of framework for access management (Allison et al., 2012).

However, much of the current services supporting e-learning require authentication, which means that the user needs to introduce his/her credentials in each one. This is a problem tackled by Single Sign-On (SSO) access control, that improves user experience and gives response to some of the issues related to identity management. With SSO systems, authentication is managed centrally and the user can navigate through different distributed applications using the same session and/or credentials (Bhatti et al., 2007). In this sense, SSO systems provide an effective way to manage authentication and authorization inside institutions, but are restricted to the administrative domain of each institution. For instance, in order to provide the mobility of users, access and exchange of contents and services across different educational institutions, the creation of federated identity management solutions is mandatory.

Federated identity management

Almost all technology-based digital identities work within particular systems and are not federated over multiple systems (Schwartz, 2010). The concept of identity federation can be understood as a group of organizations or service providers which have built trust relationships among themselves in order to enable sharing of information about the identity of its users. It returns the responsibility for authentication to a user’s home institution (also known as identity provider), and establishes authorization through the secure exchange of information between the two parties. This concept allows the possibility of integration and sharing of resources in a secure and reliable way (Bhargav-Spantzel et al., 2007).

There are a number of advantages for educational institutions in adopting a federated access management system (Kallela, 2008; Alves & Uhomoibhi, 2010). The main advantages are:

- User experience: users only need to use their institutional username and password to access internal and external resources;
- Interoperability: most of the systems are based on international standards and implemented at national level in
several countries (e.g., UK, USA, Australia, Spain, Finland):

- Cost saving: reducing the costs of service providers for setting up and managing multiple users and passwords, as well as reducing help requests from users who forget their own credentials;
- Privacy and security: is based on federation policies and trust agreements between parties, as well as standard technology, practices, and strong authorization controls over secure access channels;
- Collaboration, sharing, and resources availability: increasing the quantity, quality and variety of users, experiences and educational resources to support learning;
- Flexibility: institutions and educators can more easily track down appropriate resources for their students and negotiate licenses only for users who will use particular resources, whatever their needs and wherever they are located;
- Assessment: allowing the traceability of data related to users’ access and use of services, helping educators to adjust their instructional strategies and support to students.

In addition to e-business and e-commerce, the most common areas where these solutions are applied, federated identity management is also currently being adopted by educational institutions worldwide. These institutions have agreed in terms of policies, practices and standards in order to establish a trust relationship that allow their users to access and use services and resources provided by the partners. For example, in the European Higher Education context, eduGAIN project, from the GÉANT network, encompasses more than 40 beneficiaries, and is a service intended “[…] to enable the trustworthy exchange of information related to identity, authentication and authorisation between the GÉANT (GN3plus) Partners' federations. The eduGAIN service will deliver this through co-ordinating elements of the federations' technical infrastructure and a policy framework controlling the exchange of this information.” (European Commission Communications Networks, Content and Technology, 2013). Likewise, led by the Internet2 consortium and with more than 400 Higher Education participants, the InCommon project was created in order to support “[…] a common trust framework for U.S. education and research. This includes trustworthy shared management of access to on-line resources in support of education and research in the United States. […]” (InCommon, 2014).

Finally, there is a large diversity of identity federation systems, standards and specifications. From systems based on open standards to those based on proprietary solutions, developed by private companies. Nearly all have similar features, ranging in scope and applicability of the solution. However, some architectures and specifications stand out: the Security Assertion Markup Language (SAML - http://saml.xml.org/about-saml/) - a standard created to exchange security related information between organizations and enterprises; the Internet2 Shibboleth specification (http://shibboleth.net/about/) - an extension of the standard SAML to being an example of an identity federation system that uses Web-based implementation methodology of SAML; the Web Services Security (WS-Security, WSS: http://www.soapui.org/soap-and-WSDL/applying-ws-security.html) - which is an extension of the Simple Object Access Protocol (SOAP) developed according to the WS-Federation specification; the OpenID ( http://openid.net/) - a Single Sign-On open standard for consumers, accepted by organizations as Google, PayPal, Yahoo!, etc.; and the OAuth (http://oauth.net/about/) - another open protocol/standard that runs an API authorization between different applications; among others.

**Identified problematic scenarios and current solutions**

An extensive research agenda can be found addressing Virtual Worlds technology issues and concerns (Lee & Warren, 2007; Hendauï et al., 2008; Messinger et al., 2009; Thompson, 2011). Many areas need further exploration, but we are moving toward standardization that helps insure interoperability (Burden, 2011). However, most of the standards efforts have concentrated on the graphical aspects, neglecting other important issues that can impact its widespread adoption for education and training purposes - as identity federation.

In different dimensions of the educational process within Virtual Worlds, identity federation can impact all actors in educational institutions, such as teachers, students and staff - fostering its massive adoption. This is why the problem must be discussed by the educational technology community. In order to better explain how relevant the problem is, we present below a set of scenarios and dimensions of the educational process in Virtual Worlds, where the most cited and adopted ones (e.g., Second Life and OpenSimulator) will gain more attention - but others should also be taken into account.
Firstly, it should be possible for a teacher and/or trainer to easily create and build his/her class or team without the current administrative workload needed. This means that if a teacher desires to start running educational activities in a Virtual World, technology should automatically recognize and link their students’ academic identity with the 3D avatars, giving them the permissions and attributes based on their roles. This will enable students to access the virtual spaces with a unique set of credentials that can easily identify the student in the 3D virtual space (e.g., login, password, avatar’s name, profile information, etc.). Moreover, the usability related to this process should also be improved. For instance, this instantiation of roles and permissions can be available from a template on the web, without much complexity or difficulty to fill out. This is already happening for Virtual Worlds’ technology such as Second Life, with the SLOODLE project. However, linking avatars with the Moodle students’ mapping is needed (Sloodle, 2013).

Secondly, it should also be possible for teachers and staff to manage and track students’ attendance within the classes or training sessions. This means that teachers should have tools available to track and assure students’ participation in the classroom based on their identity, knowing when they login, logout or go outside the classroom space. Moreover, it would be also possible to have a report on avatars’ status (e.g., if they are away or not, based on the frequency of interaction with their avatar) during the class - helping manage and register the attendance and participation of each student. Clearly this is a hard task for teachers in a random access environment such as Second Life, to identify late arrivals or early departures from his/her learners, and also is a requirement of almost of the educational institutions (Madeira et al., 2010; Kluge & Riley, 2008). However, this does not fully assure the students’ participation and identity within the classes - but that is also a problem with other applications, not only with Virtual Worlds, along the years (e.g., a student can use the credential of a colleague and impersonate him; an account owner may delegate control of their attendance to a number of real life individuals who manage his/her attendance interchangeably, in order to maintain an around-the-clock in-world presence, etc.) (Lindsay et al., 2010), and even in face-to-face classes is possible to have similar problems (e.g., a student can stay in a classroom but nothing assures that he/she will not pay attention or sleep; or even in large classes with lots of students, in which is hard to manage the identities of each student during an evaluation test).

Thirdly, it should be possible to have an integrated and holistic view of students’ learning patterns, due to a transparently navigation across services, with a unique credential, that quickly can occur in a multimodal enriched learning environment (SLOODLE fails in this aspect, because it only offers the connection between Moodle and Second Life - not with the different academic services and tools, such as e-mail, repositories, videoconferencing, etc.). This means that we, as teachers or faculty members, should be able to track students’ behavior and activity based on a unique identity within all academic services and tools. Based on a comprehensive access tracking and log analysis, this would cause us to assess how students use the different resources and services in the learning process, helping to monitor it inside virtual classrooms, in order to better understand their difficulties and plan or provide adequate measures. Examples such as Moodog (Zhang et al., 2007), can be found concerning asynchronous learning supported by Content Management Systems.

Fourthly, although in a social Virtual World like Second Life having different accounts and avatars can be useful for a person (Lindsay et al., 2010), in an educational context that’s not desired. This phenomenon of identity overlapping raises important legal and ethical questions in relation to users’ rights and responsibilities in-world and ownership of avatars and their digital creations (Lindsay et al., 2010). From exploitation, to cyberbullying, sex, violence, and other problematic scenarios (Bujeja, 2007; LaChapelle, 2007), students and teachers don’t have the guarantees of security and privacy that they need. It should be possible to identify who is related to a disruptive behavior. It’s obvious that the concept of real privacy is still a utopia within Virtual Worlds, and on that point we agree with other researchers (Vilela et al., 2010; Freudenthaler, 2011), however this sense of accountability combined with access control promotes the compliance of rules and norms within an educational group or institution.

Fifthly, Virtual Worlds based on Second Life technology follow system and environment controls through Role Based Access Control (Allison et al., 2012), which, in turn, boil down to three categories: land related, content related, and group related. In each category there are different roles, permissions or privileges, and functions. It should be possible to delegate these roles automatically, based on the users’ identity and context of the learning activity. For example, in a role play activity to simulate a refugee camp in the real world, students must follow the realistic mobility options such as walking, running or crawling. If they are given the opportunity to teleport or fly, “[...] they may not intuitively experience the time taken to resolve a dilemma situation [...]” (Perera et al., 2012). Other cases, in which some engagement functions can be an obstacle to learning, such as avatar customization or
chat conversation (Kluge & Riley, 2008; Cruz et al., 2013), it should be possible to block or prohibit students to do it. For instance, if a group of architecture students have to do a collaborative task of creating and constructing a 3D virtual space, it is not desirable that they be distracted from the main focus of the activity and start to edit their avatar’s appearance. It means that they should only have access to the appropriate functions in order to complete the activity (e.g., voice, chat and build permissions).

Moreover, there are also accessibility issues related to educational scenarios in Virtual Worlds. It should be possible for the Virtual World to acknowledge if the student has any disability and adapt the 3D classroom environment for him/her (e.g., when a blind student logs in with his/her avatar in-world, the system might automatically teleport the student to the class or turn on the ‘follow’ mode connected to a colleague). Also, other students and the teacher should be aware of it. It means that if a blind student goes to a virtual class, a text-to-speech device ought to be available. Similarly, for deaf students, voice communication should only take place when a speech-into-text device is enabled. This already happens with some of the videoconferencing and Voice over IP (VoIP) technologies (e.g., Asterisk - http://zaf.github.io/asterisk-speech-recog/), where technology is flexible and capable to adapt to user’s identity and attributes (e.g., automatic translate voice into chat for a deaf user).

Sixthly, Virtual Worlds are currently mostly isolated in restricted servers, managed by a single entity (Morgado, 2013), and there is no persistent identity that other places can detect and interact with (Clark-Casey, 2010). This may change and it should be possible to optimize the distributed learning environments and resources of different academic institutions, as well as currently happens with web-based services (Morgado, 2009). It means that if a teacher in United States has a 3D science lab in their OpenSimulator server, they should be able to give access to it to an European colleague, for use with his/her students - and also the contrary, seamlessly, without requiring creation of local identities (Hypergrid technology only provides this to a limited degree, as is mentioned below). This cooperation between institutions and staff promote a network of 3D open educational resources and spaces, with new scenarios for massive learning and research. It also promotes educational mobility across countries and cultures, that can be tracked and managed based on agreed policies, practices, standards and licenses, ensuring user’s authenticity, reuse and authoring of the 3D learning objects - preventing situations like Copybot, a modified Second Life client that was able to copy illegally 3D contents (Hunt, 2007).

Finally, for research purposes, “[...] virtual worlds also add new ethical considerations to survey research. [...] identity in Second Life is fluid so ensuring respondent integrity is difficult.” (Bell et al., 2009). Thus, if we are conducting research, in most cases we should and/or may be in our interest to know who exactly is beyond the avatar, in order to ensure the validity and reliability of the data that we are collecting - preventing biases. Moreover, the user should also know who the researcher leading the study is, in order to participate in an informed and coherent way.

Ongoing projects, such as Medulla (Fox et al., 2010) or Moonshot (Howlett & Hartman, 2010), are the proof that we are pointing to the right way by stating that identity federation is an enabler for the adoption of Virtual Worlds by the educational community, if correctly applied. On the one hand, the American project Medulla, created by the Federation of American Scientists (FAS), uses web Single Sign-On access control with Shibboleth and DSpace databases manager for identity management, team building, information sharing, project management, peer review, data versioning, data archiving, intellectual property management, and learning management in Virtual Worlds (Fox et al., 2010). This clearly shows us that we are already able to use federated identities standards to access Virtual Worlds, as well as ensure that the data and metadata related to user traceability, privacy, and authoring are interoperable within different Virtual Worlds technology. However, the Medulla team had only the concern of interconnection between Virtual Worlds, neglecting web-based systems such as learning management systems.

On the other hand, the European project Moonshot, from Janet-led in partnership with the GÉANT and other partners, aims to develop a single unifying technology for extending the benefits of SAML-based federated identity to a broad range of non-web services, including cloud infrastructures, high performance computing, grid infrastructures and other commonly deployed services including email, file store, remote access and instant messaging (Howlett & Hartman, 2010). Although the Moonshot did not focus on the same concern of Medulla, this ongoing project shows us that it should be possible to communicate between different technologies using federated identity (e.g., web-based with non-web-based).
Other efforts are already being developed, but far from support the identified problems within our scenarios. For example, Hypergrid - an architecture and protocol for securely decentralizing multiuser virtual environments, establishes an open federation of multiuser applications that can exchange user agents and assets, and can generally interoperate on several basic services. It supports the teleportation of user agents between Virtual Worlds in different administrative domains while preserving user identity, as well as the users’ 3D virtual representation and connections to certain home-world services, as the users’ inventory (Lopes, 2011). However, it only assures the interconnection of different grids or servers, and has several security, scalability and trustability problems (Clark-Casey, 2010).

Moreover, projects such as the aforementioned SLOODLE (Livingstone & Kemp, 2008), simply outline how it is possible to integrate Second Life Virtual World and Moodle. One of the most fundamental affordances of it is to pair Moodle users to their Virtual World avatars. When a user clicks on the Second Life registration booth, while logged in with their avatar, they are prompted to visit a Moodle registration page. This allows Moodle to verify the Second Life identity of the Moodle user, and this data is then stored in Moodle. Alternatively, a 'Login Zone' object in Second Life allows avatar registration to be driven from Moodle, followed by logging into Second Life. More similar efforts are found in the literature (Madeira et al., 2010).

Future directions

As our main goal, we have shown before a set of clear scenarios in which identity federation solutions can bring an added value, namely: classroom management, content reuse, learning analytics, accessibility, research, etc. On the one hand, identity federation can simplify the process of creation, management and monitoring of classes in 3D Virtual Worlds, by improving the usability, reducing the time and the administrative workload involved, and, in turn, the frustration and unpleasantness this may cause to educators. On the other hand, by ensuring private and secure access based on students’ roles/attributes, identity federation can help somewhat to prevent surface learning, or even to guarantee the same rights and conditions to students’ learning. Moreover, it can reduce and/or eliminate issues related to other technological challenges of Virtual Worlds, as cost or content production. If trust relations between institutions exist, within the use of Virtual Worlds, we are only one step behind the concept of 3D open education - by reusing shared resources and spaces, within institutional copyright policies and accountability terms.

However, the way we think and put the issue of federation must go beyond the interconnection between different Virtual Worlds using the same underlying technology (Lopes, 2011), or specific learning management systems (Livingstone & Kemp, 2008). We need to take e-learning, in a near future, as a unique shared service, independently from the platform or application in use (e.g., web or grid-based). We argue that the standardization of Virtual Worlds within themselves, without concerns of identity insurance, user traceability, privacy, accountability or interoperability with other non-virtual world systems can affect seriously the adoption of these for learning purposes. The path to interoperability and a seamless interaction between information systems, learning systems, third party applications and even devices, point out that the focus should be in the user, where the achievement and management of the correct identity is crucial to this success. The correct identity management of users, complemented with security and trust aspects, are intrinsic factors of identity federation standards, therefore should be taken into account into the technological development and standardization efforts of Virtual Worlds. As consequence of this, its adoption in academic or even non-academic learning spaces can be greatly enhanced.

We clearly need to interconnect digital identities with physical identities, thus creating a unique federated identity system that can act independently from the Virtual World in use. Instead of the need to adapt Virtual Worlds to the federation systems, we need an open agreement with basic federated and interoperable standards that should be capable to adapt within the broad range of available services on the Internet. This way, Virtual Worlds are able to scale and evolve independently within the different research areas, without compromising the users’ identity. Our proposal remains on the convergence of the Medulla (Fox et al., 2010) and Moonshoot (Howlett & Hartman, 2010) federation efforts that can be achieved by the development of new technological solutions or readapting the current ones. As seen before, we know that many educational institutions worldwide have agreed to the adoption of the same policies, practices and standards that allow users’ authentication and authorization while protecting their privacy. In this sense, we can straightforwardly reach and simplify the widespread adoption of Virtual Worlds by the educational community worldwide: implementing a unified solution for cross-compatible federations, that can not only federate and manage identities between 3D Virtual Worlds, but also within different systems and applications (independently from the platform in use).
Acknowledgements

This work is funded (or part-funded) by the ERDF – European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness) and by National Funds through the FCT – Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project «FCOMP - 01-0124-FEDER-022701». This paper is based on an invited presentation under the SLACTIONS workshop on “Technology Challenges of VWs in Education & Training toward widespread adoption,” at the 5th International Conference on Games and VWs for Serious Applications, VS-Games 2013, held in Bournemouth, United Kingdom, 11-13 September 2013.

References


Technical Problems Experienced in the Transformation of Virtual Worlds into an Education Environment and Coping Strategies

Murat Coban¹, Turkan Karakus²*, Asiye Karaman², Fatma Gunay² and Yuksel Goktas²
¹Agri Ibrahim Cecen University, Vocational School, Computer Technologies and Programming, Agri, Turkey //
²Ataturk University, K. K. Faculty of Education, Computer Education and Instructional Technology, Erzurum, Turkey //
mcoban2005@gmail.com // turkan.karakus@gmail.com // asiye.karaman.89@gmail.com //
fatmagunay9@hotmail.com // yuksel.goktas@hotmail.com
*Corresponding author

ABSTRACT
Research into virtual worlds demonstrates that to successfully use virtual world platforms in different disciplines, certain limitations and potential difficulties of these platforms must be overcome. The current study extends previous research by investigating problems in integrating virtual worlds in education with a longitudinal observation of two cases and in perspectives of designers, practitioners and administrators. The purpose of this study was both to determine technical problems encountered by practitioners and designers who create environments in 3D virtual worlds for a certain purpose and to suggest practical strategies for these problems. In the study, designers experimented with two different 3D virtual world platforms for one year, to design and create learning environments. To collect the research data, interview records and design diaries in which the designers and practitioners noted problems that they encountered during two virtual world projects. The results indicate that the technical problems has not aroused only related the problems and those problems influenced several pedagogical issues like motivation of students and guidance in the environments. Our findings should be beneficial both to virtual world designers and to users who wish to use these environments in broad range of fields.

Keywords
Virtual worlds, Second Life, Open Simulator, Material design

Introduction
Virtual worlds are three-dimensional (3D) virtual environments wherein users come together and communicate via the Internet (Gamage, Tretiakov, & Crump, 2011). Virtual world environments allow designers to create fantastic objects; architects can build online construction projects and participate in cooperative work; researchers can conduct experimental studies; and designers can create various realistic environments and use tools to facilitate social experiences (Hai-Jew, 2010). Virtual world platforms also support cooperative education and allow educators to prepare simulations for distant education. This helps students to gain social experience, to develop prototypes, and to visualize scientific facts (Rufer-Bach, 2009).

In recent years, a number of educators have focused on the influence of increasingly-popular 3D virtual worlds on teaching and learning (Choi & Baek, 2011). All educational institutions want to invest educational opportunities which are really beneficial for learning. Virtual world platforms requires a considerable cost to be implemented and thus pedagogical benefits are supposed to be high. However, now many educators’ options have been narrowed since Second Life removed its educational discount (Yilmaz, Reisoglu, Topu, Karakus & Goktas, in press). Before implementing a successful teaching and learning in virtual worlds, the limitations and potential difficulties involved in the use of virtual worlds must be clarified (Peachey, Gillen, Livingstone, & Smith, 2010). If such problems can not be solved, then it may not be possible to determine the real influence of these environments on students. Potential learning time might also be wasted, due to technical issues (Childs, Schnieders, & Williams, 2012). Solutions to these problems will allow virtual worlds to become useful environments, which anyone can access from any place. Globally, they might then accommodate millions of digital users (Hinrichs, Hill, & Patterson, 2011). This study exemplifies what kind of problems might be encountered during the implementation of educational virtual worlds starting the establishment phase to end user implementation phase. Besides, study will show how those problems were solved to provide a pedagogically effective learning environment.
Related studies

In virtual worlds, specific objects to be used for learning can be designed and developed. However, designing and coding are two basic problems related to virtual worlds (Galea, Legarreta, Martí, Gisbert, Rallo, García, & Cela, 2011). As Duncan, Miller and Jiang (2012) stated technological and usability issues are key area of current and future research in virtual worlds. Using the Open Simulator (OpenSim) platform, Schmeil (2012) reported that it was difficult for users to design their own avatars, and that it was not possible at all to overcome problems experienced while deactivating certain features of the environment. In another study, Dickey (2005) used two different virtual world platforms to assess the appropriateness and limitations of their design. Dickey reported that the two virtual worlds either had limitations or provided designers with different facilities. For instance, while one of the virtual world platforms helped users to store 3D objects, the other virtual world platform allowed users to make use of more developed control options for the design and development of 3D objects and environments. Maddrell, Watson, and Morrison (2013) used virtual worlds for problem-based engineering applications and realized that the development platform used to create a model in the design process was time-consuming to employ; that the tools used in the design process should be easier to use; and that for this purpose, designers should be provided with technological support. Bronack, Cheney, Riedl, and Tashner (2008) opined that users should only need to expend minimal effort in designing virtual worlds, and that the tools used for communication should be technically supported. Smelik, Tutenel, Kraker, and Bidarra (2011) reported that it was not an easy process for many designers to design content in virtual worlds, and that the tools used for communication should be technically supported.

Several research studies investigate the problems related using virtual worlds in educational settings. González, Santos, Vargas, Martín-Gutiérrez and Orihuela (2013) investigate a university-wide project in perspectives of teachers and students. They revealed technical issues and general satisfaction with virtual worlds. Warburton (2009)’s study handles one of the virtual world platform, Second Life, designer experiences showed that there are several challenges and advantages in terms of technical, immersive and social aspects. Kluge and Riley (2008)’s study also reveals some technical and institutional (cost, etc) issues related virtual worlds. Some of previous research also focused on one or a few issues and especially technical problems of virtual worlds (Pfeil, Ang & Zaphiris, 2009; Thackray, Good & Howland, 2010). However, literature lacks the studies investigating the challenges encountered starting the selection of the design platform to the presentation of designed environment to end user. While studies focuses on a few aspects of challenges using virtual worlds in education, they came up with a fact that each challenge in using virtual worlds requires the careful consideration of a number of design possibilities (Warburton, 2009). This study might illuminate design possibilities with a wide scale examination of different implementations and with the perspectives of all stakeholders of educational virtual world implementations.

Virtual worlds have many affordances like creating new learning places, interaction, immersion and virtual representation (Dickey, 2011). As Dickey (2011) stated these affordances differ for each virtual world environment. And these affordances can only be opportunities as long as they are served without problems. Virtual worlds used for education should be as free as possible of technological difficulties and should impose minimal limitations on both designers and users. Technical problems might not be solved only by virtual world platform servers but also they might be overcome by designers with different pedagogical strategies. The goal of this case study was to discover problems that were experienced by designers and users of 3D virtual worlds, and to collect their suggestions to solve these problems. The study does not claim that it offers any breakthrough approach but it differs from previous research in that it handles two longitudinal educational virtual world implementation cases where different platforms used, collects data from all stakeholders of the projects, reveals a broad range of problems in different phases of the projects and suggest practical solutions for each problem. The results are important in terms of presenting problems from setting the virtual world platform up to use the developed platform for target learners. The results of this study will be helpful to students, educators, and designers who use 3D virtual worlds, and to software developers who wish to develop the functions of these environments. The following research questions guided this study:
• What technical problems are experienced by the designers while they are selecting the virtual world platform, and what solutions do they suggest for these problems?
• What technical problems are experienced by the designers during the design and development phases of a virtual world environment, and what are their suggested solutions for these problems?
• What technical problems are experienced by the users during the implementation process of a virtual world environment, and what are their suggested solutions for these problems?

Method

This study examined the development processes of two environments in the virtual world platforms OpenSim and Second Life (SL), which were used for two different projects. OpenSim is one of the most important 3D platforms that support virtual learning (Campbell, Wang, Hsu, Duffy, & Wolf, 2010). OpenSim is an open-source and multi-user virtual world platform, in which users can privatize virtual worlds for different purposes. The platform permits clients to access environments from any location and to input designs as required. It also uses a 3D interface (Konstantinidis, Tsiatsos, Demetriadis, & Pomportsis, 2010). OpenSim provides service via its support of open-source code, which is input through servers controlled by users (Ryoo, Techatassanasoontorn, Lee, & Lothian, 2011).

In the first project examined in this study, the designers planned and developed 3D objects in an OpenSim environment as part of the course Project Development and Management, which they took in the Fall term of the academic year 2011-2012 at the Department of Computer Education and Instructional Technology in a large-scale university located in East Anatolian region of Turkey. In this course, over a period of six months the designers developed 3D environments to introduce European Union projects at TUBITAK (The Scientific and Technological Research Council of Turkey), and then they created presentations which were related to these programs in the OpenSim environment. The purpose of the OpenSim project was to help pre-service teachers to develop 3D materials with the use of virtual platforms. A Facebook group was formed to allow the designers to share their 3D designs, information, and the problems that they faced during the project development process. Figure 1 presents a screenshot of a sample 3D design created by some of the designers for this OpenSim project.

In the second project, the Second Life (SL) virtual world platform was used to generate a greater awareness of winter sports. SL is a platform that allows designers to develop 3D objects, and to store and reuse interactive objects (Girvan, Tangney, & Savage, 2013). Unlike an OpenSim environment, the SL platform requires a fee to use many of its features and does not have an open-source code structure. In 1999, an SL 3D environment created by Linden Lab in San Fransisco provided thousands of users the opportunity to come together simultaneously (Linden Lab, 2013). Users of a SL platform can buy objects to meet their social needs in the environment, or they can sell objects that they produce themselves to other users (Shelton, 2010).
In the SL platform, the purpose of the winter sports project developed in 2012-2013 was to raise elementary school students’ awareness of and interest in winter sports such as Ice Hockey, Curling, Speed Skating, and Artistic Skating by using the SL environment. The students had the opportunity to learn each of the sports activities in the selected winter games via multimedia tools such as graphics, videos, animations, and educational games in the SL virtual world environment. The designers noted in their design diaries all of their experiences throughout this process, from the development of this environment to its application. Figure 2 shows a user studying with a 3D application in the SL environment.

Figure 2. A sample application area in the SL environment

### The Participants

Table 2 presents the researchers’ roles in the two different projects regarding the processes of analysis, design, development, and implementation. In both projects, the researchers did their best to supervise the participants, to coordinate the activities carried out in the 3D environments, and to meet the participants’ expectations regarding the technical, application, and equipment-related problems regarding the 3D virtual environment. The project executor and the technical support team for the OpenSim project also worked on the winter sports project in SL. All of the research data in this study were based on the experiences of the team described in Table 2 below.

<table>
<thead>
<tr>
<th>Environment</th>
<th>Number of Participants</th>
<th>Role</th>
<th>Job</th>
<th>Work experience in virtual worlds</th>
</tr>
</thead>
<tbody>
<tr>
<td>OpenSim</td>
<td>1</td>
<td>Project leader</td>
<td>Faculty member</td>
<td>Three years</td>
</tr>
<tr>
<td></td>
<td>42</td>
<td>Designer and Developer</td>
<td>Senior students</td>
<td>One-year experience in design</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Technical support</td>
<td>Research assistants</td>
<td>Two years</td>
</tr>
<tr>
<td>Second Life</td>
<td>1</td>
<td>Project executor</td>
<td>Faculty member</td>
<td>Two years</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Supervisor</td>
<td>Faculty members</td>
<td>Eight years</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Technical Support</td>
<td>Research assistants</td>
<td>Two years</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Interface designer</td>
<td>Masters students</td>
<td>Two years</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Practitioners</td>
<td>Research assistants</td>
<td>Two years</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Material developer</td>
<td>Masters students</td>
<td>Two years</td>
</tr>
</tbody>
</table>
The data collection tools and data collection

This study examined problems that the designers of two different projects experienced during the processes of analysis, design, development, and implementation. The designers’ interview records were used as data for the OpenSim Project; and designers’ design diaries were used as data for the SL Project. For SL project the designers kept the design diaries for about one year. The duration of time spent from the analysis phase to the implementation phase was about six months for the OpenSim project and about one year for the SL Project. Table 3 shows the processes followed in the OpenSim and SL projects.

Table 3. The processes followed in the two projects

<table>
<thead>
<tr>
<th>Months</th>
<th>OpenSim Project</th>
<th>Winter Games project in SL</th>
</tr>
</thead>
</table>
| 1st month | • OpenSim was introduced to the students. It was made ready for use, and the students were divided into study groups.  
• The Facebook communication group was formed. | • Information about the virtual world to be rented for the project was assessed, and the island was rented.  
• An appropriate virtual platform was selected for the activities.  
• The island was divided into certain study areas, and the design process was begun. |
| 2nd-4th months | • The groups were informed about the subjects for which they would create the designs, and 3D design areas were formed for each group for the development of the materials.  
• Feedback was provided for the studies. | • In order to enrich the designed areas with multimedia elements, draft studies were begun, and experts were asked for their support. |
| 5th month | • The study groups made their presentations.  
• The problems encountered and related solutions were discussed. | • For the applications, animations for each winter sports area were designed. |
| 6th month | • Focus group interviews were held with the designers.  
• The Facebook activities of the designers were examined. | • The problems experienced in the processes of development of the interactive audios, graphics, videos, and animations were overcome.  
• The problems encountered during the arrangements were determined and noted, and proposed solutions were examined. |
| 7th-12th months | | • In order to allow the users in the environment to benefit from the oral communication activities, necessary test processes were conducted.  
• Before the applications, the infrastructures of the schools were examined.  
• For the pilot studies, the necessary preparations were made by the technical team in each school, and the computers were made ready for use. This process took approximately one week for each school. |

For the analysis of the research data collected for the winter games project in SL, the meeting records and the design diaries kept by all the designers and the development team were used. For the OpenSim project, transcriptions of the data collected from the designers’ focus group interview records were analyzed using the content analysis method. Of the 42 designers who worked in the OpenSim project, 36 participated in the interview process. Regarding the
OpenSim project, the data collected from the Facebook records were categorized and arranged according to the research questions, and related themes were derived. Lastly, the problems revealed during the two projects were categorized; repetitive problems were eliminated; and all of the problems encountered were presented.

Findings

The findings show that five theme categories of problems were experienced regarding the design of virtual world environments. These categories, as can be seen in Figure 3, were the end users, the environment in which the application was conducted, the design decisions (by the designer), the platform in which the design was done, and the physical environment in which the platform was created.

![Image of Figure 3: Problem types encountered during the 3D virtual environment’s design and application]

As can be seen in Figure 1, all of the categories featured certain problems during the initial formation of the environments, and during the design processes and applications. Among the five categories, the physical environment in which the platform would be established, the platform to be designed, and the design decisions constituted the most significant problems. Regarding the implementation process, the learning environment designed, the context in which the implementation was done, and the end users’ abilities characterized the problems that occurred during the application. The overall findings obtained from the two projects were gathered and summarized in Table 4. These problems are categorized and described under headings.

<table>
<thead>
<tr>
<th>Problems encountered during the formation of the environments</th>
<th>Problems</th>
<th>Environments</th>
</tr>
</thead>
<tbody>
<tr>
<td>• OpenSim server configuration problems</td>
<td></td>
<td>OS</td>
</tr>
<tr>
<td>• Long, complicated, and time-consuming procedures related to the purchasing process to make the SL environment ready for designs</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>• Transferring the multimedia elements onto the platforms and freezing problems</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>• Lack of technical and equipment features necessary to form the environments</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>• Inefficient network infrastructures and Internet band width</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>• Obligatory coding of the oral communication add-ons</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Problems encountered during</th>
<th>Problems</th>
<th>Environments</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Problems experienced by the designers particularly regarding sources and information in the native language concerning programming.</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>
the design and development of the environments technical issues, and design

- Problems related to programming and design due to lack of facilities provided by the environments ✓ ✓
- Web add-ons designed for the users failed to work, and some media elements could be used by more than one user, which hinders the work of other users. ✓
- Problems experienced with the processing and use of audio files due to the limitations of the environments ✓
- Recovering objects deleted accidentally or deliberately from the 3D environment ✓ ✓
- It is difficult and sometimes impossible to establish interactions between the 3D objects and physical simulations. ✓ ✓
- Current 3D objects necessary for each project are few in number, and sometimes they fail to meet the needs of the users. ✓ ✓
- Failure to back up the environments or 3D objects developed in the SL environment ✓
- Failure to control the objects due to the coordinates of the 3D objects, or it is difficult to do so ✓ ✓
- Limited duration of audios (10 seconds) hampers the design ✓
- Sometimes, objects freeze and are locked-up. ✓ ✓
- Dialogue boxes constantly open to ask for the user’s permission to activate the animations. ✓
- It is difficult to direct a user to a website when an interactive object is clicked, and the environments do not simply allow the use of such web processes. ✓
- Informative 3D objects fail to attract the users’ attention. ✓
- The users lose their direction or fail to navigate in the related areas. ✓ ✓
- The users have difficulty putting on or changing their clothes on their avatars using the interactive objects or fail to do it correctly, and dialogue boxes which constantly open for this purpose influence their motivation negatively. ✓
- Failure to monitor the users’ behavior automatically and to provide feedback ✓

Problems encountered during application of the environments

- Problems related to oral communications between users ✓ ✓
- Failure to control the audio tools when more than one user joins the same environment, which causes overlaps of voices ✓
- The users lose their direction in the environment and fail to find their direction. ✓
- The users fail to use the input devices effectively to control their avatars. ✓ ✓
- Motivation problems experienced by the students in the process of adaptation to the environments ✓
- Locking-up and freezing problems experienced regarding the use of animations and other interactive objects ✓ ✓
- The users establish oral or written communication with each other rather than fulfilling the tasks assigned to them in the environment, and failure to restrict the communication tools hinders the application process. ✓
- The users skip the interactive objects, videos, and animations in the environment, do not conduct the related applications, and do not fulfill the tasks assigned to them. ✓
- The users choose inappropriate avatars, as the avatar options provided by the SL environment cannot be controlled. ✓
Problems encountered during the formation of the virtual world environments and proposed solutions

The most important infrastructure elements that are required during the formation of environments are server computers and an Internet connection. In order for the designers to access the environments and to design, server computers must have the necessary technical features and equipment. However, the technical features of current server computers are likely to cause data loss for designers. Concerning the platforms that require fees, certain problems will likely be experienced during the purchasing process. For example, in order to make the SL environment ready to design, purchasing an individual island did not pose a problem, but the complex and costly financial procedures for institutions to purchase an island (e.g., the obligation to purchase five islands at a time) caused problems during the formation of the environments. Table 5 summarizes these problems and their causes.

<table>
<thead>
<tr>
<th>Problem in the environment</th>
<th>Its cause</th>
<th>Solution to the problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problems related to installing the 3D virtual world software on the computers</td>
<td>Inefficient equipment in the computers used</td>
<td>New equipment was purchased to meet the needs of the SL program, and the necessary additional software was installed.</td>
</tr>
<tr>
<td>The time-consuming purchasing process for the SL environment</td>
<td>Complex and costly procedures for institutional renting</td>
<td>Purchasing of islands with an individual-personal account</td>
</tr>
<tr>
<td>Waste of time due non pre-structuring of the software, hardware, and other sources necessary to access the environments</td>
<td>The need to adapt the Internet network configurations and servers (necessary for the environments) to virtual worlds</td>
<td>Adjusting the necessary settings and installing the related add-ons.</td>
</tr>
</tbody>
</table>

Problems encountered during the design and development phases and proposed solutions

The design and development of interactive objects in virtual worlds in particular requires knowledge about programming, technical issues, and design. In order to design and develop environments, designers need to use the programming languages supported by the environments and to develop their designs accordingly. Second Life uses the Linden Scripting Language (LSL). The developers need to spent a considerable time to solve particular problems in controlling the animations, programming the sound effects, and using the programming language necessary for the animations to work under predetermined conditions.

Although the basic purpose of developing 3D environments is not to prepare simulations or animations, simulations can be prepared when current programming and construction tools are effectively used. However, it took a lot of time to test the physical collisions (between objects), and to establish interactions between the avatars and physical objects and between the objects created. Table 6 summarizes not only the problems that arose due to the limitations of the environments but also the causes of these problems.

<table>
<thead>
<tr>
<th>Problem in the environment</th>
<th>Its cause</th>
<th>Suggested solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access to sources is limited.</td>
<td>Some project based needs, requires very sophisticated codes</td>
<td>Experienced teams were asked for help using the social media and other communication tools, especially from those designed a sport related environment</td>
</tr>
<tr>
<td>It was not easy to develop the desired interactive objects.</td>
<td>There was limited built-in software to facilitate the development of the interactive objects, or it was difficult to use the provided software.</td>
<td>Using the tools provided by the SL interface, 3D objects were designed, and no additional program was used, design decisions changed.</td>
</tr>
<tr>
<td>The web add-ons led to conflicts due to common use.</td>
<td>The web viewers in the environments showed only one address at a time, and all the users saw this.</td>
<td>The web pages were allowed to be opened in Explorer, except for SL.</td>
</tr>
</tbody>
</table>
Failure to transfer some of the audio files into the environment | The SL environment supported only 10-second audio files. | Uploading audios longer than 10 seconds on YouTube, and a related link was included in the desired area.

Dialogue boxes constantly opened to ask for users’ approval to activate the animations. | For safety purposes, SL was programmed in a way to ask constantly for users’ permissions. | No solution was found, yet the users were provided with intensive guidance to help them avoid making mistakes.

Inability to retrieve the objects accidentally or deliberately deleted in the environment | Lack of functions in the environment to recycle the deleted objects | The objects sent to the recycle bin were recovered piece by piece.

Failure to establish interactions between the objects (for instance, in Ice Hockey, it was not possible to move the stick and puck according to the collision velocity when the stick hit the puck) | It was not possible to program real-life simulations in the platforms. | No solution was found, animations were simplified or performed by button click.

It was not easy to backup the objects developed in the SL environment. | The SL software did not have an open-source code structure, and the current backup software was not effectively used. | Backing up was done using third-party software.

There were problems experienced in controlling the interactive objects. | The object coordinates were confusing. | No solution was found. There was a waste of time in the design process.

Problems encountered during the implementations of the 3D virtual worlds and proposed solutions

The problems encountered during the applications were due to the provided equipment, to the design of the environments, to the students’ skills, and to the technical infrastructure of the environments. A slow Internet connection was one of the most significant problems during the application. The purpose of design in a 3D environment might identify the limitations that might be faced. In the Second Life study, as the researchers’ purpose was to study psychomotor learning, they had to overcome a number of technically difficult situations. One of these difficulties was due to the presentation of all the required information in the learning environment. It was difficult to make the intensively designed information attractive to the students. The texts on the objects were not legible due to the user’s viewing angle; they therefore did not draw the users’ attention; and these objects consequently did not help the users to carry out tasks assigned in the environment. In addition, while in the environment, the students were expected to progress step by step; but the guidance elements did not generally work. The students wanted to navigate freely in the environment. The texts that informed the users about how to do the required movements did not arouse their interest. So they skipped these movements without applying them. Some communication problems revealed because the oral communication options supported by the environment were not efficient. When more than one person spoke at the same time in the environment, overlapping of voices occurred. Table 7 summarizes the problems encountered by the practitioners in the environments as well as the causes of these problems.

<table>
<thead>
<tr>
<th>Problem in the environment</th>
<th>Its cause</th>
<th>Suggested solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral communication between the users was difficult.</td>
<td>When the Internet speed was slow, and overlapping of voices occurred.</td>
<td>The amount of written communication in the environment was increased.</td>
</tr>
<tr>
<td>As it was possible for the students to choose non-human avatars, they were unable to demonstrate some of the behavior required by the environment.</td>
<td>It was not possible to restrict the avatar options.</td>
<td>The avatars were restricted by controlling students face-to-face during the implementation.</td>
</tr>
<tr>
<td>When the audio objects were clicked simultaneously, overlapping of voices occurred.</td>
<td>There were no efficiently developed tools to control the objects for oral interaction in the environment.</td>
<td>The audio directions were removed, and interactive information objects were used.</td>
</tr>
</tbody>
</table>
The users failed to find their direction in the environments. | The add-ons and built-in effects failed to develop the direction tools in a way that would arouse the users’ interest and help to guide them. | Flying in the environment was blocked to prevent the users from losing their direction, and the users were allowed to move by walking and running as well as by teleportation. |
---|---|---|
It was difficult to read the informative 3D objects. | It was not possible to provide users with efficient components that added effects and other striking elements to the 3D objects, and some of the information was not visible to the users. | Figures and colors to draw the students’ attention were used (colored footprints, information objects in the shape of a snowman, etc.). |
The students lacked motivation during the process of adaptation to the environment. | Especially concerning use of the 3D animations found in the environments, dialogue boxes constantly opened to ask for the user’s permission. | No solution was found to this problem. The users were provided with intensive guidance to help them avoid making mistakes. |
Problems of locking-up and freezing were encountered regarding the animations and other objects. | When the Internet band width used was not sufficient and with more than 25 users (SL was supposed to support 100 users and Open Sim supports 25 users at a time) | A limited number of students were allowed to enter the environment at the same time. |
The students focused on communications when they were expected to perform specific tasks. | The SL interface software was not flexible enough to deactivate all the text-based or oral communications options or to allow conducting these types of communications with the involvement of an administrator. | The students were provided with lists of activities and asked to do them. The students completing the activities joined the social area and established written or oral communications with their friends freely. |
The multimedia tools were not effectively used. | There was a lack of add-ons, or the Internet infrastructure used was not sufficient. | In cases of a lack of infrastructure, a new temporary computer laboratory was established at the university, and the students were transferred to these environments. |
The users experienced difficulty putting on and changing the avatars’ clothes by using the interactive objects found in the environment, or they failed to do so appropriately. | The dialogue boxes constantly opened in the process of wearing any object in the environments, and there were permission-related problems. | The clothes were designed as objects by the designers, and the users were allowed to put on all the clothes at the same time with a single click. (However, in this case, the user had to do certain other things to take off the clothes.) |

**Discussion and conclusion**

The designers who developed the 3D materials in the virtual world platforms of OpenSim and SL pointed out that limitations caused by the platforms and also by the end-users, practitioners, and designers who developed the 3D environments may have influenced the overall experience. While designing the platforms, the type of platform used, the selected features of the physical environment, the design decisions, the utility of the technical-equipment in the environments, and the end-users’ knowledge, skills, and expectations all may have all been the sources of problems. Such problems as uploading errors, and locking-up and freezing when working with add-ons like graphics, audios, videos, and animations generally resulted from the limitations of the platforms. Design decisions, were related to the platforms were also likely the sources of problems. For example, in the SL environment, the users’ ability to select or change their avatars as they wish is not normally considered a problem, but it can be a problem if the avatars are meant to be only a certain type to fit the design. SL does not allow a restriction of options for changing the avatar. Regarding the problems experienced in the application process, these problems seem to have occurred during the process of the users’ adaptation to the platforms, and also because of their computer use skills and the features of the environment. All of the problems that were specifically related to the platforms are presented in Table 8.
Table 8. Potential problems in the environments

<table>
<thead>
<tr>
<th>Platform</th>
<th>Potential problems that may be encountered in the environments</th>
</tr>
</thead>
<tbody>
<tr>
<td>OpenSim</td>
<td>• There are server problems.</td>
</tr>
<tr>
<td></td>
<td>• The computers have inefficient technical features, and this results in the problem of freezing.</td>
</tr>
<tr>
<td></td>
<td>• It is not possible to retrieve the objects deleted in the environment.</td>
</tr>
<tr>
<td></td>
<td>• It is not easy to activate the oral communication options.</td>
</tr>
<tr>
<td></td>
<td>• It is not possible to download the objects from the server, or it takes time to do so.</td>
</tr>
<tr>
<td></td>
<td>• It is not easy to activate the multimedia add-ons.</td>
</tr>
<tr>
<td></td>
<td>• The interface software used to access the environment does not allow easy designing of objects.</td>
</tr>
<tr>
<td></td>
<td>• There are not enough sources appropriate to the native language regarding the software.</td>
</tr>
<tr>
<td></td>
<td>• The design process takes a lot of time and requires effort.</td>
</tr>
<tr>
<td></td>
<td>• The coordinates of the objects are confusing and make the design process difficult.</td>
</tr>
<tr>
<td></td>
<td>• It is not possible to find enough technical information or to receive help.</td>
</tr>
<tr>
<td></td>
<td>• The audio files in the 3D environment support the users for only 10 seconds at most.</td>
</tr>
<tr>
<td></td>
<td>• It is very difficult to develop interactions between 3D objects or the physical simulations.</td>
</tr>
<tr>
<td></td>
<td>• Users making designs and developments in the 3D environments complain about the complexity of the interfaces used to access the environment, and there are motivational problems due to the users’ expectations and desire make the design process easier.</td>
</tr>
<tr>
<td></td>
<td>• It is not possible to retrieve the objects deleted in the environment.</td>
</tr>
<tr>
<td></td>
<td>• More than one person can use the same media elements simultaneously.</td>
</tr>
<tr>
<td></td>
<td>• Dialogue boxes constantly open to ask for approval to activate the animations.</td>
</tr>
<tr>
<td></td>
<td>• Users have difficulty putting on, purchasing, and changing the clothes for their avatars when using the interactive objects found in the environment, or they fail to do so appropriately.</td>
</tr>
<tr>
<td></td>
<td>• The web add-ons designed for the users fail to work well, and they allow more than one person to use the same media elements simultaneously.</td>
</tr>
<tr>
<td></td>
<td>• It is difficult to control the interface with the input devices.</td>
</tr>
</tbody>
</table>

Second Life

<table>
<thead>
<tr>
<th>Potential problems that may be encountered in the environments</th>
</tr>
</thead>
<tbody>
<tr>
<td>• The procedures for purchasing an island in the SL environment are long, complex, and time-consuming.</td>
</tr>
<tr>
<td>• The computer equipment must possess the required technical features so that the environment can work.</td>
</tr>
<tr>
<td>• There are not enough sources in Turkish related to the platform.</td>
</tr>
<tr>
<td>• The development process for the multimedia objects takes a long time, and it is difficult to control the objects.</td>
</tr>
<tr>
<td>• The avatar options provided by the environment are not appropriate to the educational content.</td>
</tr>
<tr>
<td>• The audio files in the 3D environment support the users for only 10 seconds at most.</td>
</tr>
<tr>
<td>• It is not easy to develop interactions between the 3D objects or physical simulations.</td>
</tr>
<tr>
<td>• There are problems with authority and permissions.</td>
</tr>
<tr>
<td>• Users making designs and developments in the 3D environments complain about the complexity of the interfaces used to access the environment, and there are motivational problems due to the users’ expectations and desire make the design process easier.</td>
</tr>
<tr>
<td>• It is not possible to retrieve the objects deleted in the environment.</td>
</tr>
<tr>
<td>• More than one person can use the same media elements simultaneously.</td>
</tr>
<tr>
<td>• Dialogue boxes constantly open to ask for approval to activate the animations.</td>
</tr>
<tr>
<td>• Users have difficulty putting on, purchasing, and changing the clothes for their avatars when using the interactive objects found in the environment, or they fail to do so appropriately.</td>
</tr>
<tr>
<td>• The web add-ons designed for the users fail to work well, and they allow more than one person to use the same media elements simultaneously.</td>
</tr>
<tr>
<td>• It is difficult to control the interface with the input devices.</td>
</tr>
</tbody>
</table>

According to the data from the two projects, both platforms had a number of technical limitations. Therefore, in order to transform these 3D virtual worlds into real learning environments, it is necessary to improve their technical aspects (Ibáñez, García, Galán, Maroto, Morillo, & Kloos, 2011). When pedagogically considered, our findings demonstrate that the users lost their motivation due to technical and equipment-related problems and when they need to access certain functions through a process that involves several steps caused them to skip these functions. Therefore, educators carrying out applications in 3D virtual worlds should consider these potential problems and certainly prepare a ‘Plan B’. However, renovations and changes made in the development processes of projects are likely to have negative influences on both the productivity of the project team and the productivity of those who conduct applications in the environment. For this reason, educators interested in virtual environments need to gain confidence and competence regarding the interruption of pedagogical activities due to technical problems (Bower, Cram, & Groom, 2010). It may require time to overcome problems regarding permissions, or regarding the accessibility of objects or the design of the environments. Thus, educators should possess a number of skills, so that they can create effective designs for virtual worlds and cope with such situations (Warburton, 2009).
Three-dimensional environments will not be attractive for educational purposes if their technical obstacles cannot be removed. Many researchers assume that the influence of virtual worlds on teaching and learning soon will become global, but they note that the current use of these environments for educational purposes is still limited (Jarmon, Traphagan, Mayrath, & Trivedi, 2009). Lack of understanding the behavior of users in virtual worlds, technological problems and the cost of environments, health problems experienced by users who depend on their computers, and adaptation requirements for environments are reported to be among the biggest limitations to the use of virtual environments (Eschenbrenner, Nah, & Siau, 2008). In order to improve both 3D environments and interface software functions, features must be added which will better facilitate the design and development processes for objects in environments. Add-ons should be included that will facilitate the programming process to develop environments. Add-ons can also be used to overcome the usability problems reported in this study. Rather than considering only the factor of cost in assessing the effective use of virtual technologies in education, instructional designers should design and develop instructional materials more carefully for these environments (Merchant, Goetz, Cifuentes, Keeney-Kennicutt, & Davis, 2014). The following suggestions are additionally offered for consideration:

- Prior to the application process, the Internet bandwidth and the related network infrastructure should be assessed, especially in underdeveloped and developing countries. This could help to overcome the problems of locking-up and freezing in the environments.
- In order for 3D virtual platforms to work effectively, the technical and equipment-related features could be kept at as low a level as possible.
- Regarding the use and programming of 3D virtual environments, sources in different languages could be developed for designers and users.
- Future studies could focus on the extent to which technical problems lead to pedagogical problems.

Acknowledgments

This study was conducted as part of the “Effects of Virtual and Multimedia Environments on Interest and Awareness Towards Winter Sports” project (number “111K516”), supported by the The Scientific and Technological Research Council of Turkey Project.

References


Childs, M., Schnieders, H. L., & Williams, G. (2012). "This above all: to thine own self be true": Ethical considerations and risks in conducting higher education learning activities in the virtual world SL (TM). Interactive Learning Environments, 20(3), 253-269.


Distributed Pervasive Worlds: The Case of Exergames

Teemu H. Laine1* and Carolina Islas Sedano2

1Department of Information and Computer Engineering, Ajou University, Woncheon-dong, Yeongtong-gu, 443-749 Suwon, South Korea // 2School of Computing, University of Eastern Finland, P.O. Box 111, 80101 Joensuu, Finland // tlaine@ajou.ac.kr // carolina.islas@uef.fi

*Corresponding author

ABSTRACT

Pervasive worlds are computing environments where a virtual world converges with the physical world through context-aware technologies such as sensors. In pervasive worlds, technology is distributed among entities that may be distributed geographically. We explore the concept, possibilities, and challenges of distributed pervasive worlds in a case study—an exergame entitled Running Othello. Combining a board game with sensor-enhanced physical activity, Running Othello aims at motivating young players to be physically active. Based on the analyzed literature and mixed-method evaluations of Running Othello, the following contributions emerged: (i) a total of 13 design principles for distributed pervasive exergames, (ii) the players’ perceptions with distributed gameplay, (iii) an analysis of development challenges, and (iv) four dimensions of distribution in Running Othello. Player evaluations were conducted on university students and on children and young adults participating at a science festival. These results are useful scholars and developers interested in the concept of pervasive worlds. With this exploratory paper, we have taken the first steps toward comprehending the conceptualization, design, implementation, and evaluation of distributed pervasive worlds.

Keywords

Pervasive world, Game, Exercise, Distribution, Context-awareness

Introduction

Virtual worlds, such as Second Life, typically run on desktop computers. Although mobile applications are available for virtual worlds (e.g., Mobile Grid Client, Lumiya, Pocket Metaverse), they often ignore the smartphone’s capacity to adapt to the user’s experience by detecting the context. One of the current challenges in the field of education is to enable context-sensitive learning experiences (JISC, 2006). To support educators using alternative approaches to learning—such as behaviorism, cognitivism, and constructivism—computer scientists should explore, develop, and understand technical systems that aim to facilitate the achievement of objectives set by educators and students (JISC, 2006). Therefore, to overcome the technological hurdles facing virtual worlds in education, we set out to explore the possibilities that exist beyond the computer-simulated virtual worlds by considering the converging space between the physical and the digital world. Pervasive world is a system that enables the interplay between physical and digital worlds by detecting the user’s context. Technologies that implement a pervasive world, or any other pervasive system, include smartphones, sensors, wireless communication devices, and algorithms that enable context-awareness (Chalmers, 2011). This technology can link resources from the physical world to the virtual world, allowing novel interactions, such as using the body itself as a controller (Kilili & Merilampi, 2012).

Milgram and Kishino (1994) proposed a reality-virtuality continuum that encompasses the real environment, augmented reality, augmented virtuality, and the virtual environment. We expanded the reality-virtuality continuum to include world types that utilize these environments (Figure 1). Virtual worlds are based on virtual environments, while pervasive worlds may utilize augmented reality to enhance the visualization of, and interaction with, the virtual content. In augmented reality, real-world images (e.g., from a camera feed) are augmented with virtual elements. By contrast, augmented virtuality enhances the virtual-world by embedding real-world resources (e.g., from a web-camera feed). We see pervasive worlds as the next generation of virtual worlds—releasing users from their keyboards to explore the physical world with context-sensitive digital content.

In virtual and pervasive worlds, technical components and users are distributed over a network. The notion of distribution is important, from both a technical and conceptual perspective, for understanding the challenges and possibilities for virtual and pervasive worlds. This paper is among the first to explore the concept of distributed pervasive worlds, with a focus on their design, implementation, and deployment. Educators and scholars interested in pervasive worlds will benefit from our study.

As a case study for distributed pervasive worlds, we present Running Othello—a distributed pervasive exergame (DPE) combining a board game with physical exercise. An exergame was chosen as the target of our exploration in part because obesity...
is a global challenge (World Health Organization, 2014), and games are intrinsically motivating and immersive (Garris, Ahlers, & Driskell, 2002; Islas Sedano, Leendertz, Vinni, Sutinen, & Ellis, 2013; Malone & Lepper, 1987; Sweetser & Wyeth, 2005).

By reviewing the literature and analyzing the development and evaluation processes in Running Othello, we aim to address the following questions: What aspects are important in the design process of DPEs? How do the players perceive the concept and distributed gameplay in Running Othello? What challenges does developing a DPE? What are the dimensions of distribution for Running Othello? Before investigating Running Othello, we survey the literature to understand DPEs generally and to determine their design principles.

**Literature survey**

**Exergames**

Exergames combine physical exercise with digital gaming to deliver cognitive and physical challenges in an attractive package—with physical, psychological, cognitive, and academic benefits (Staiano & Calvert, 2011). Modern exergames utilize context-aware technology, which is also used in pervasive games (Magerkurth, Cheok, Mandryk, & Nilsen, 2005).

Researchers have coupled exergames and networking to enable distributed gameplay. *Table tennis for three* (Mueller & Gibbs, 2007) combines a table tennis board with video conferencing to allow three players to play against each other on distributed installations. The players are able to communicate through video and audio channels. *Virtual Network Marathon* (VNM) players participate in a marathon over a network connection (Zhang et al., 2012). The players run on sensor-enhanced treadmills as they observe their avatars in a virtual 3D environment while learning about Chinese culture and the Beijing Olympics. In *Virtual Tug-of-War* (Harfield, Jormanainen, & Shujau, 2009) players of two distributed teams pull a physical rope connected to a device measuring the force of pull. This measuring device relays the force data to software that visualizes the data and exchanges it with software running on the opposing team.

Smartphones eliminate the need for specialized controllers and enable outdoor gameplay in exergames. In *Zombies. Run!* (Six to Start, 2012), a player is chased by virtual zombies, and the smartphone’s sensors track the player’s location, motion, and speed. *GeoBoids* (Lindeman et al., 2012) uses augmented reality, GPS, and audio to allow the player to interact with and capture GeoBoid creatures by running after them. *Health Defender* (Wylie & Coulton, 2008) is a Space Invaders clone where the player shoots down viruses. A heart-rate monitor is used in bonus tasks. Kiili and Merilampi (2012) use the smartphone’s accelerometer for simple exergames, monitoring physical activities such as squats and jumps.

The design for most exergames focuses on physical activity, such as running (VNM, Zombies. Run!), or pulling (Virtual Tug-of-War). What if we could focus on strategy in games, with physical activity as a consequence of gameplay? To explore this possibility, we transformed the strategic board game, Othello, into an exergame.

**Design principles for distributed pervasive exergames**

In ideal circumstances, design principles guide the application development process. In the case of novel systems, such design principles are often derived after analyzing empirical interventions. In developing Running Othello, we
did not have any design principles for DPEs to guide us. However, we derived design principles from exergame literature to aid in the analysis and discussion of our study, and these can later serve as guidance for developers.

The VNM was built using the ISCAL (Immersion, Scientificalness, Competitiveness, Adaptability, Learning) design model (Zhang et al., 2012). An exergame should support immersion such that the players feel that they are participating in a realistic sporting experience. Scientificalness implies that the game’s exercises are scientifically valid. Competitiveness calls for competitors to be included to increase the feeling of a competitive sport. An exergame should also be adaptable to the players’ respective fitness levels. Finally, the game should include learning content.

The Dual-Flow model (Sinclair, Hingston, & Masek, 2007) was developed for optimizing the effectiveness of exergames. The model (Figure 2) explains how a player can enter the flow in two dimensions: attractiveness and effectiveness. For an exergame to be attractive, a balance should be reached between skill and challenge. To be effective, it must have an optimal intensity-to-fitness ratio. Because all these properties are subjective, the exergame should be adaptive to player profiles.

Gao and Mandryk (2011) proposed the concept of casual exergames for motivating people to exercise in small chunks of time. They demonstrated the concept with a PC-based exergame, using the Microsoft Kinect motion detector, and discovered nine design principles for casual interaction and efficacious exercise (Figure 3). Gao and Mandryk derived the design principles from casual-games research and based efficacious exercise principles on the recommendation that effective exercise must continue for a period of 30 min per day where the heart rate and caloric expenditure reaches a certain level.
Social exergames comprise social interaction via online gameplay. The ExerLink platform (Park et al., 2012) supports the construction of social exergames that can be played with heterogeneous exercise devices such as jumping ropes, treadmills, and stationary bicycles. The ExerLink platform design was based on (1) the adaptation to exercise- and player-specific differences, (2) adaptability to network latency, (3) the provision of rich interaction modalities, and (4) the provision for exercise history information.

We identified 13 design principles for DPEs from the reviewed literature (Figure 4), grouping them into five topics organized into three layers. The Technical layer contains essential principles for any reusable pervasive system development (Baldauf, Dustdar, & Rosenberg, 2007; Laine, Islas Sedano, Sutinen, & Joy, 2010). The Socio-cognitive layer focuses on the user’s mind and interactions. It comprises principles that aim at making the system appealing and useful through social and cognitive dimensions. Finally, the Exergame layer contains principles related to exergames.

<table>
<thead>
<tr>
<th>Physical exercise</th>
<th>Social interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Physical challenges</strong> (Gao &amp; Mandryk, 2011; Sinclair et al., 2007; Zhang et al., 2012): an exergame should offer versatile, scientifically valid physical challenges at different levels to maintain the game's effectiveness.</td>
<td><strong>Social interaction</strong> (Park et al., 2012; Sweeters &amp; Wyeth, 2005): an exergame should provide means for social gameplay and interaction. Social interaction is an important social motivator together with competition and collaboration.</td>
</tr>
<tr>
<td><strong>All ages and fitness levels</strong> (Gao &amp; Mandryk, 2011): Game content should be designed to support exercise for any ages from child to senior regardless of their fitness levels.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Physical exercise</th>
<th>Gaming</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clear rules and controls</strong> (Gao &amp; Mandryk, 2011; Garris et al., 2002; Malone &amp; Lepper, 1987): an exergame should have clear rules and simple controls that motivate the player instead of confusing them. If the game has unfamiliar rules or controls, a tutorial might be necessary.</td>
<td><strong>Cognitive challenges</strong> (Sinclair et al., 2007; Zhang et al., 2012): an exergame should offer cognitive challenges such as strategic thinking, problem solving and knowledge acquisition. Without challenging the player's mind an exergame can feel dull.</td>
</tr>
<tr>
<td><strong>Immersion</strong> (Islas Sedano et al., 2013; Malone &amp; Lepper, 1997; Sinclair et al., 2007; Sweeters &amp; Wyeth, 2005; Zhang et al., 2012): gameplay should be immersive in order to take the player into the flow where they are completely focused on the game experience. Immersion can be facilitated by various game motivators such as curiosity, challenge, fantasy, competition and collaboration.</td>
<td><strong>Raising awareness</strong> (Zhang et al., 2012): an exergame should have facilities to raise awareness on a topic that connects to the context in which the game is played or to the context of exercise and sport. This way the game becomes a pedagogical asset. Awareness can be raised by quizzes, bulletins and other microlearning activities (Hug, 2007).</td>
</tr>
</tbody>
</table>

**Figure 4. Design principles for DPEs**

These principles are a starting point for the design and analysis of DPEs. Although they are described exclusively from the DPE perspective, the technical and socio-cognitive principles are generic and thus applicable to any distributed pervasive world. For example, Adaptability might refer a system’s ability to adapt its behavior to learner profiles and learning contexts. Furthermore, the Gaming principles are pertinent to any game-like pervasive world.

**Running Othello**

“It was fun and different. You can see that a mobile phone doesn't always have to have a negative effect on health.” (F16)
The game concept

Othello is a strategic turn-based board game where two players compete on an 8×8 board using black and white pawns. All pawns have the same properties and available moves depending on their position on the board. The players capture the opponent's pawns by moving their own pawn over it in a straight line (Figure 5). The game ends when the board is full, when there is only one color left, or when no further moves are possible.

Running Othello is an Android-based exergame version of Othello, developed for the SciFest 2013 science festival in Finland. Players in Korea and Finland played together in a virtual world using their respective physical environments. The Running Othello board size can be customized to any physical dimensions and grid size to encourage running. For example, a 10×10 board might be deployed in a large area in a park where local matches can take place on a single board. However, Running Othello encourages distributed gameplay over a network connection on boards.
distributed worldwide. Near-Field Communication (NFC) tags are used as game-board cells for quick interaction. Numbers on the NFC tags and on the virtual board are identical, so the player must visually match the boards. To make a move (i.e., to conquer a cell), the player touches a tag with a smartphone initiating a random mission. If the mission is successful, the smartphone displays an updated virtual game board. There are three mission types: physical, awareness, and reaction (Figure 6). An awareness mission is a quiz on the importance of water, in accordance with SciFest’s theme.

As a design decision, we removed turn-taking from the game to encourage running and thinking quickly, resulting in two consequences. Firstly, sporty and clever players can make consecutive moves. We used additional pawns in the initial setup to prevent a skilled player from finishing the game too quickly. Secondly, two players can read the same tag simultaneously. In such cases, the first player to successfully finish the mission conquers the cell.

**Distributed implementation**

Distributed gameplay is enabled with a server running on J2EE Servlets to which game clients connect via HTTP. The system architecture is presented in Figure 7. The server manages user profiles, channels, the board/rules, and the database. After first configuring the server address in client’s preferences, the player must register a new profile. Then, the player can login and create a game channel or join an existing channel. A single channel supports two players, but the server can handle multiple channels. Once a channel is full, the match begins. Two players can share the same physical board or play on distributed boards. One board can host multiple simultaneous matches. In every case, the board must be constructed from the NFC tags before playing. Tags must be programmed with numbers indicating their location on the board. The game client requires Android 2.3 or higher. Apart from tags and smartphones connected to the server, no other equipment is needed for gameplay.

![Figure 7. Architecture for Running Othello](image)

The content of a read tag is sent to the board/rule manager for validation. If the move is invalid, an error is returned. Otherwise, the server returns a randomly chosen mission, or—if the player is lucky—the cell is conquered directly.
The player must complete a given mission before the cell is conquered. Clients receive game-board updates from the server after conquering the cell or by polling.

Extensibility is important in any learning system for facilitating reusability. Running Othello’s extensibility is twofold. Firstly, new quizzes can be added to the database without updating the software. Secondly, missions are implemented as Android Activities that are linked to the game-board Activity. New missions (e.g., educational tasks) can be added to the system with simple source-code modification. When the player makes a move (i.e., reads a tag), any mission attached to the game can be launched. This way, Running Othello can be used as a platform for any type of game activity.

Evaluation

To identify the technological hurdles facing distributed pervasive worlds we evaluated Running Othello from two perspectives:

- Players’ perceptions of gameplay, answering the research question: How do the players perceive the concept and distributed gameplay in Running Othello?
- Developers’ perceptions on the game-development process, answering the research question: What challenges does developing a DPE entail?

The gameplay was evaluated in three sessions with a total of 28 players. After local evaluations in Korea during two sessions, we evaluated the game between players in Korea and Finland at SciFest 2013 where players communicated over Skype. Figure 8 illustrates the evaluation settings with distributed gameplay.

Figure 8. Evaluation settings

Figure 9 indicates the process for a mixed-method evaluation to answer the research questions above. The gameplay data were collected with a questionnaire and through observations. The questionnaire was based on a previous exergame evaluation (Kim et al., 2014). The questionnaire included demographics, qualitative and quantitative (Likert scale) instruments on previous gaming and exercise experiences, positive and negative experiences with Running Othello, motivation, game features, usability, suitability, and a comparison with computer games and sports. The questionnaire was available in Korean, Finnish, and English. The development evaluation utilized the data from developers’ essays written after the project.

For the evaluations, we first installed the game on Samsung Galaxy S2 and S3 smartphones that connected to the server in Korea. The players were first presented with the game concept and research objectives. The players then played against opponents for approximately 2-10 min during which they only used Running Othello and no other features of the phone. During the gameplay, researchers and developers noted observations, and prompted the players for verbal feedback after playing. Finally, the players answered the questionnaire and were permitted to ask for clarifications.
Gameplay evaluation

The following presents the evaluation results on features, ease of use, and suitability of Running Othello with mean ($\mu$) and standard deviation ($\sigma$) values, followed by a correlation analysis. Statements that were exclusive to the SciFest questionnaire are marked “SF.”

Features

Statements 1-5 in Figure 10 refer to the question “Which of the following features were interesting?” Among the features, using NFC tags was rated the highest (93%) whereas Othello game rules received the lowest rating (68%), possibly because 25% had no opinion and 29% reported that they did not know Othello beforehand. These results followed these open question answers regarding what they liked or enjoyed about the game:

“Touching the tag is amazing and fun. It is a fun game.” (M13).
“It is fun to think to get more points than my partner to win.” (F21).
“A real opponent. A feeling of doing something vs. sitting on a computer.” (M33).

Figure 9. Data collection process

Figure 10. Game features (1 = Othello game rules, 2 = Using NFC tags, 3 = Competing against another player, 4 = Solving quizzes (SF), 5 = Solving movement challenges (SF), 6 = Motivation: Running, 7 = Motivation: Using the mobile phone)
The aforementioned comments reveal some motivators (curiosity about technology, competition, social gameplay, and hands-on experience) complemented by the developers’ perceptions (curiosity, novelty, and exercise):

“The game fascinated the kids somehow since they can play with Korean players, use smartphones, and explore surprising tasks. For example, one kid tried to spin as fast as possible. Consequently, he fell down, but he still laughed loudly. In aspects of curiosity and technology, the game succeeded.” (M24).

“Two of the teachers showed great interest in the game. They believed it is a new type of game and will make kids do more exercise when playing.” (F29).

Statements 6 and 7 answer the question “What was the reason for you to continue playing (what motivated you)?”, thus measuring how running and phone usage influenced the players’ motivation. Running connects to aforementioned hands-on experience and exercise motivators, whereas phone usage relates to curiosity about technology, and particularly to interactions through NFC and sensors, which bore novelty in the target contexts at the time of evaluation. The phone usage motivated most players (71%) while many had no opinion regarding running (39%), thus leaving positive answers for 36%. This can be partly explained by the test locations where limited board sizes did not require much running.

**Ease of use**

Statement 1 measures a phone’s easiness as an interaction tool for DPEs (Figure 11). In contrast to traditional user-phone-interaction, interaction in Running Othello occurs via embedded motion sensors and NFC tags on the floor. The result of statement 1 (79%) shows that the novel way of using the phone was considered as easy, yet it could be improved (see comments below). Understanding the challenges (89%) and reading tags (78%) were also regarded as easy. The ease of game’s controls received 68% positive and 25% negative answers. Poor sensitivity in the sensors might explain some of the negative answers. Players reported this and other problems, when asked what they disliked or found difficult in the game:

“Some tasks didn't work properly. For example, in turning I had to spin 10 times to count.” (M25).
“Finding point [the NFC tag] is difficult.” (M13).
“I didn’t know the game rules the first time.” (F21).

![Figure 11. Ease of use (1 = Using phone was easy for me, 2 = Understanding challenges was easy (SF), 3 = Reading tags was easy (SF), 4 = Controlling the game was easy)](image)

A developer suggested that Othello’s unfamiliarity could be mitigated were the rules better explained:

“Many kids do not know this game before; they felt it is hard to get the rules. However, after we presented [the game] they were getting excited.” (F29).
Suitability

The game’s suitability to the players’ skill levels, available time, and money is reported in Figure 12. It was considered easy for beginners (79%), and most thought it matched their skill levels (86%). Statements 3 and 4 were both rejected by 46% of the players, while 36% gave no opinion. Statement 5 was answered by 19 players, most of whom were willing to pay up to $4.99. This shows that Running Othello could go beyond prototyping.

![Figure 12. Suitability (1 = This is easy game for beginners, 2 = This game is suitable for my skill level, 3 = Game was too long, 4 = Game was too short, 5 = How much would you pay for the game)](image)

The game can feel long due to repetitive missions. One player suggested that the game should include more variety when we asked how it might be improved:

“Don’t make the same mission appear. Add more variety of missions.” (M25)

Correlations

Table 1 presents statements having significant correlations (<=-0.5, >= 0.5). Age (1 = 13-17, 2 = 18-22, 3 = 23-26, 4 = 27-30, 5=>30) correlates negatively with running as a game motivator (-0.57) and with the game being easy for beginners (-0.5). This suggests that young players enjoyed the running more than older players and that they felt the game was more approachable.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Age</td>
<td>-</td>
<td>-0.49</td>
<td>-0.19</td>
<td>-0.57</td>
<td>-0.4</td>
<td>-0.5</td>
<td>-0.42</td>
</tr>
<tr>
<td>2. Interesting feature: Othello game rules</td>
<td>-</td>
<td>-</td>
<td>0.2</td>
<td><strong>0.53</strong></td>
<td>0.05</td>
<td><strong>0.65</strong></td>
<td><strong>0.76</strong></td>
</tr>
<tr>
<td>3. Motivation: using mobile phone</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.31</td>
<td><strong>0.5</strong></td>
<td>0.23</td>
<td>0.22</td>
</tr>
<tr>
<td>4. Motivation: running</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.16</td>
<td>0.35</td>
<td>0.34</td>
</tr>
<tr>
<td>5. Willingness to pay for the game</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.07</td>
<td>-0.3</td>
</tr>
<tr>
<td>6. Easy game for beginners</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td><strong>0.72</strong></td>
</tr>
<tr>
<td>7. Suitable for my skill level</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

The simplicity of Othello’s rules made it easy for beginners (0.65) and at a suitable skill level (0.76), two qualities that also correlated (0.72). Furthermore, those who liked Othello’s rules were also often motivated by running (0.53).
The players who were motivated to play with a mobile phone were more willing to pay for the game (0.5). The culture of purchasing mobile applications among children and young adults may have an effect on these results, and this should be explored in future research.

**Development evaluation**

In 2012 fall two Korean undergraduate computer science students created the first prototype for Running Othello. In January 2013, one Vietnamese and two Chinese students from a Finnish university joined the project. Korean students spent three weeks in Finland to initiate the distributed work. When the project was completed, the developers wrote essays on their experiences. The analysis of the data, including our observations, resulted in 11 challenges and solution proposals (Table 2) which can help distributed pervasive world developers.

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Explanation</th>
<th>Solution proposals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Network connectivity</td>
<td>In a distributed environment, the network is often a bottleneck. During SciFest, the game server was located in Korea, causing a noticeable delay in client-server communication. The game was playable, but location transparency failed.</td>
<td>Slow connection speeds can be mitigated by providing adequate network resources between clients and servers, by minimizing the number of server calls, and by applying data compression.</td>
</tr>
<tr>
<td>2. Activity detection</td>
<td>In exergames, detecting the player’s activity and physical status is essential. Some players reported that the game did not react properly to physical activity. There are two possible reasons for this: (1) The sensor data was poorly processed, and (2) the players were performing the missions incorrectly. For example, the activity detection algorithm can fail if the phone is held incorrectly.</td>
<td>Some physical missions used the deprecated orientation sensor. To improve the results, missions compute orientation from the accelerometer and magnetic field sensors. Low-pass filtering could be used to remove noise from the sensor data. Players should also be instructed to perform the missions correctly.</td>
</tr>
<tr>
<td>3. Distributed development</td>
<td>Distributed game development between two countries was challenging in terms of time and cultural differences. Because Korean students started the project earlier, it was difficult for the students in Finland to join, thus limiting their work input. The most productive time for collaboration was when two Korean students travelled to Finland.</td>
<td>With any global software development, special care must be taken to foster multicultural collaboration. All members should join the project as early as possible. Furthermore, student-researcher exchanges should be encouraged and cultural differences mitigated by the project leader.</td>
</tr>
<tr>
<td>4. Contextualization</td>
<td>Running Othello was designed to be portable across contexts. The game rules are easy, and the content was translated into target languages. Despite this, one of the challenges was to integrate the game with SciFest’s water theme.</td>
<td>A successful pervasive exergame should serve the target context’s needs. The user interface and content should be customizable so that the game can be adapted to different contexts. In Running Othello, we used customizable quiz missions to match SciFest’s theme.</td>
</tr>
<tr>
<td>5. Multiplayer management</td>
<td>To run multiple concurrent matches, a multi-channel mechanism was developed. This was not part of the original prototype server design.</td>
<td>Server design should focus on multiplayer aspects. Furthermore, support for multiple concurrent game sessions should be planned from the...</td>
</tr>
</tbody>
</table>
### Discussion

The players evaluated Running Othello’s features, ease of use, and suitability. The game’s technology played a central role in the players’ perceptions. Positive ratings given for the NFC tags usage may have been the result of...
their novelty because NFC tags were uncommon technology at the time of evaluations. From the perspective of ease of use, one interesting point is that despite the difficulty that some players experienced in locating tags, reading them was considered easy (78%). These results suggest that the NFCs are suitable for pervasive exergames, but mappings between virtual and physical worlds must be clearly defined.

Because of the fact that gameplay is short, it is not possible to comprehensively analyze the game’s immersive aspects. However, there is evidence that most players were stimulated by the challenge in terms of mastering the technology, competing and solving the puzzles (Figure 10). One developer cited curiosity as a motivator. It is possible that the novelty of the technology, or the game system itself, raised the curiosity of the players.

Some players complained that mission instructions were unclear and activity detection was poor. We used graphical illustrations to show players how to perform the physical missions, but these may not have been sufficiently clear. A tutorial could be used to rectify this, and the quality of the sensor data could be improved with filtering. These could increase the success rate of missions and player satisfaction.

Although suitability statements showed positive results, it is difficult to say whether the game would remain suitable given long-term deployment. Strong correlations between Othello’s rules, its ease for beginners, and its suitability to varying skill levels suggest that basing the game on Othello without turn-taking was a successful design.

Many players answered quantitative statements with “No opinion”—e.g., regarding the “Othello game rules” (25%), and whether the “Game was too long/short” (36%). Perhaps these statements were poorly understood by some players. The questionnaire should be improved in order to increase its clarity and encourage respondents to have an opinion.

To facilitate an understanding of the meaning and implications of the evaluation results, we identify four dimensions of distribution in the following section. Additionally, we analyze the applicability of our 13 DPE design principles for Running Othello. Finally, we discuss the educational potential for Running Othello and the limitations of this research.

**Dimensions of distribution**

Through the analysis of the game’s concept, development, and evaluation, four dimensions of distribution emerged: focus, gameplay, technology, and development.

**Distributed focus** refers to a situation where the user’s attention is shared between multiple activities. In Running Othello, the player’s focus is distributed between activities that occur simultaneously in the virtual world (the virtual game-board, awareness and reaction missions, and the virtual opponent) and the physical world (the physical game-board, physical missions, and a real opponent). The player must sustain cognitive thinking and strategic reasoning under time pressure and physical exertion. Supported by Gameplay Evaluation: Features; Development Challenge 7: Game-board visualization.

**Distributed gameplay** has been studied from several perspectives. Massively Multiplayer Online Role-Playing Game (MMORPG) researchers have investigated how virtual worlds can be handled technically by dividing them into regions across servers (Assiotis & Tzanov, 2006). The rise of pervasive technology has called for explorations beyond technical challenges into game mechanics that combine distributed gameplay and enjoyment together (Mueller & Gibbs, 2007). In Running Othello, distributed gameplay manifests itself as a real-time battle between players in different locations. By providing player-to-player competition and communication, the feeling of togetherness can be enhanced. Unlike MMORPGs, it is not necessary to divide computing across multiple servers because there are few concurrent players. Supported by Gameplay Evaluation: Features, and Development Challenge 10: Player communication.

**Distributed technology** enables distributed gameplay over the internet. Conceptually, Running Othello’s client-server architecture enables players from anywhere in the world to play against each other. Practically, some players are affected by limitations such as communication latency and cost. Designing distributed technology requires special attention to the implementation of efficient communication and to graceful error handling. Supported by
Development Challenge 1: Network connectivity; Challenge 5: Multiplayer management; Challenge 6: Game-board updates & synchronization.

The distributed development for Running Othello was conducted by students in South Korea and Finland. The developers gained unique experience from this international collaboration, allowing them to work remotely and also face-to-face. The distributed development provided a natural environment for distributed evaluation. Supported by Development Challenge 3: Distributed development.

Running Othello and design principles

We used the 13 proposed design principles for DPEs (see Figure 4) as a tool for analysis. Table 3 exemplifies the applicability of each design principle for Running Othello with supporting data gathered in this study.

<table>
<thead>
<tr>
<th>Principle</th>
<th>Applicability in Running Othello</th>
<th>Supported by</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical exercise</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All ages and fitness levels</td>
<td>Game rules and content suit everyone. Physical missions should be scientifically valid for all age groups and fitness levels.</td>
<td>Gameplay Evaluation: Suitability, Challenge 11: Player motivation</td>
</tr>
<tr>
<td>Physical challenges</td>
<td>Running Othello has physical missions, but they should be scientifically valid. Moreover, physical challenges that are more effective should be created.</td>
<td>Gameplay Evaluation: Features, Suitability, Challenge 2: Activity detection</td>
</tr>
<tr>
<td>Gaming</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clear rules and controls</td>
<td>Othello’s rules are well-known and simple. Controls for the game are easy, but additional instructions are needed for physical missions.</td>
<td>Gameplay Evaluation: Usability, Suitability, Challenge 7: Game-board visualization, Challenge 9: User interface, Challenge 8: Multimedia, Challenge 11: Player motivation</td>
</tr>
<tr>
<td>Immersion</td>
<td>Challenge is a motivator in strategic games. Running Othello is motivating in the regard, as well as with regard to curiosity. Other motivators (e.g., fantasy) could be included in future research.</td>
<td>Gameplay Evaluation: Features, Challenge 8: Multimedia, Challenge 11: Player motivation</td>
</tr>
<tr>
<td>Sociality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Competition, collaboration</td>
<td>In Running Othello, two players compete against each other. Currently there are no collaborative tasks.</td>
<td>Gameplay Evaluation: Features, Challenge 5: Multiplayer management, Challenge 10: Player communication</td>
</tr>
<tr>
<td>Social interaction</td>
<td>Playing against a human rather than a computer was a key design decision. The interaction with a remote opponent could be improved by text, voice, and video chat features.</td>
<td>Gameplay Evaluation: Features, Challenge 5: Multiplayer management, Challenge 10: Player communication</td>
</tr>
<tr>
<td>Cognition and learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive challenges</td>
<td>The player’s cognition is stimulated with Othello’s strategic challenges and the demand for visual coordination. The quiz presents additional cognitive problems to solve.</td>
<td>Gameplay Evaluation: Features, Challenge 7: Game-board visualization</td>
</tr>
<tr>
<td>Raising awareness</td>
<td>Running Othello included a quiz to raise awareness on the importance of water. In the future, the game could also raise awareness about exercising and the player’s body.</td>
<td>Challenge 4: Contextualization</td>
</tr>
<tr>
<td>Distributed pervasive architecture</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adaptability</td>
<td>Running Othello is not adaptable to different user profiles or surrounding contexts.</td>
<td>Challenge 4: Contextualization</td>
</tr>
</tbody>
</table>
Context-awareness is achieved with the NFCs and smartphone sensors. This can be improved by enhancing data processing and using body sensors (e.g., a heart-rate monitor).

Extensibility
It is easy to launch any Android Activities (i.e., missions) from Running Othello. Quiz missions can be extended with customized questions.

Portability
Transferring Running Othello between contexts is easy because it does not depend on the physical location. The game is deployable anywhere with the basic technology (smartphones, networking, and tags).

Distributed operation
The architecture of Running Othello was designed to support distributed gameplay. Communication delays should be minimized.

- Gameplay Evaluation: Features
- Challenge 2: Activity detection
- Challenge 4: Contextualization
- Challenge 6: Game-board updates and synchronization
- Gameplay Evaluation: Features, Suitability
- Evaluations locally in Korea, and globally between Korea and Finland
- Challenge 1: Network connectivity
- Challenge 5: Multiplayer management
- Challenge 6: Game-board updates and synchronization
- Challenge 10: Player communication

The proposed design principles applied here as a tool for analysis constitute the first means for guidance in comprehending the complexity of DPEs. These principles are not intended to replace existing design principles for games, human-computer interaction, or pervasive software development. Rather, they are intended to complement previous knowledge, and to offer guidelines for critical areas typical to DPEs.

Educational potential of Running Othello

It is noteworthy that Running Othello is more than a DPE. It functions as a distributed pervasive world platform that can be extended with missions (i.e., with Android Activities). When a player makes a move, any mission (e.g., an educational task) can be launched. Running Othello has potential to support the following:

- The **learning objectives** of the educator to use the platform as a tool for gamification or to support game-based learning and facilitate the immersion and motivation of students.
- The **learning activities** of students to achieve the educator’s vision: Cognitive challenges related to a specific theme (e.g., illustrate sport injuries, ask mathematical questions, or show language flash-cards). A suitable pedagogical model for this is so-called *microlearning*, where a large topic is divided into micro-learning units to be studied in brief sessions (Hug, 2007)
  Physical challenges that entice the player to learn about their physical capabilities (e.g., in physical education classes).
- A **contextualized impact** on the learner by involving contextual resources with collaboration and competition. Developers can utilize context data (e.g., sensors and the weather) to make missions relevant to the student’s particular context.
- **Novel interactions** enabled by the distribution dimensions of the system that are not otherwise possible (e.g., in real-time global play with local objects).

Limitations

The study is limited as follows: (1) The sample size (28) was too small to make generalizations or claim a high degree of reliability, (2) the play time was too short to identify the long-term impacts, (3) the physical missions were not validated with scientific guidelines, and (4) the educational potential of Running Othello was not verified. These limitations will be addressed in future research.
Conclusions

The concept of pervasive worlds is an evolutionary shift away from virtual worlds toward the border area straddling the physical and virtual world. This process is facilitated by the ubiquity of smartphones and their capacity to deliver context-sensitive experiences to the user. To answer four research questions, we presented the development and evaluation of the Running Othello exergame as a case of distributed pervasive worlds.

The research question “What aspects are important in the design process of DPEs?” was answered with a literature review. As the result, 13 design principles for DPEs emerged. The gameplay evaluation answered the question “How do the players perceive the concept and distributed gameplay of Running Othello?,” suggesting that the concept of Running Othello has the potential to motivate and support players to explore and discover various subjects beyond their physical locations through play. However, the results also indicated that the merging of physical and virtual worlds introduces diverse challenges. To answer the question “What challenges does developing a DPE entail?,” we analyzed developers’ reflections and our own experiences with the development challenges. Finally, based on multiple analyses of the evaluation results and the development process, four dimensions of distribution (viz., focus, gameplay, technology, and development) emerged to answer the question “What are the dimensions of distribution for Running Othello?”

The dimensions of distribution we identified introduce many technical and conceptual challenges that need to be addressed in the development process. Some technical challenges were presented in this paper, but a complete taxonomy remains to be created. The proposed design principles can alleviate some of the challenges, but we have just now taken the first steps toward a thorough comprehension of the conceptualization, design, implementation, and evaluation of distributed pervasive worlds.

Acknowledgements

We are grateful to Hochul Yang, Kwangsu Kook, Shawn Xiaolei, Chung Dao, Zhao Yang, Seugnet Bliignaut, and Mike Joy. This research was supported by the New Faculty Research Fund at Ajou University (2012).

References


What is the Relationship between Technology and Mathematics Teaching Anxiety?

Enver Tatar¹*, Yılmaz Zengin² and Türkan Berrin Kağızmanlı¹

¹Department of Secondary Science and Mathematics Education, Atatürk University, Turkey // ²Department of Secondary Science and Mathematics Education, Dicle University, Turkey // entatar@gmail.com // yilmazzengin@outlook.com // turkanberrin@hotmail.com

*Corresponding author

(Submitted February 17, 2014; Revised May 26, 2014; Accepted July 1, 2014)

ABSTRACT

The aim of this study is to determine the relationship between pre-service teachers’ perceptions regarding technology use in mathematics teaching and their computer literacy levels as well as their mathematics teaching anxiety. The nonexperimental correlational research, which is included in the quantitative research approach, was used in the study. A total of 481 pre-service mathematics teachers constitute the sample of the study. The mathematics teaching anxiety scale, a perception scale for technology use in mathematics teaching, and the computer literacy scale were used as data collection tools. Based on the analysis of the obtained data, a low-level, negative and significant relationship was found between pre-service teachers’ mathematics teaching anxiety and their perceptions regarding technology use in mathematics teaching. Also a low-level, negative and significant relationship was found between pre-service teachers’ mathematics teaching anxiety and their computer literacy levels. It may be concluded that pre-service teachers’ mathematics teaching anxiety decreases as their perception levels regarding technology use in mathematics teaching positively increases and their computer literacy levels increase.

Keywords

Mathematics teaching anxiety, Computer literacy, Pre-service teachers, Perceptions regarding technology use in mathematics teaching

Introduction

Mathematics anxiety emerges when students exhibit illogical emotional reactions or when they are expected or required to solve mathematics problems (Aydin, 2011). It is considered among the common affective factors that have drawn attention in recent years (Baloğlu & Koçak, 2006). Many researchers have tried to define mathematics anxiety. Among these researchers, Dreger and Aiken (1957) defined mathematics anxiety as a syndrome of affective reactions exhibited towards mathematics whereas Ashcraft and Faust (1994) defined it as the tension, helplessness and mental disorganization that individuals experience when they are required to perform an operation with numbers and figures or when they solve a mathematics problem. Mathematics anxiety may cause the emergence of such behaviors in students as avoiding mathematics courses and developing negative attitudes towards activities that require mathematical operations (Ma & Xu, 2004). Mathematics achievement, mathematics performance and applications to education are the factors that impact on anxiety (AsHcraft & Krause, 2007). According to Gresham (2010), mathematics anxiety is defined as both an affective and a cognitive feature in the nature of individuals who experience learning problems. Since it is a frequently encountered condition in every stage of education, it is important to understand and define, and to avoid or reduce mathematics anxiety. According to Skemp, memorized rules and the manipulation of symbols with little or no meaning are more difficult to learn than an integrated conceptual structure, and this can result in major slipping blocks for the student (as cited in Newstead, 1998). The very abstract nature of mathematical symbols adds to the difficulties that people encounter when learning mathematics, including difficulties in storing and using information in working memory (AsHcraft & Krause, 2007). It was found by researchers that anxiety towards mathematics plays a significant role in students’ failure in this course (Ari, Savaş, & Konca, 2010; Baloğlu, 2001; Cates & Rhymer, 2003; Ma & Xu, 2004; Minato & Yanase, 1984; Satake & Amato, 1995; Sherman & Wither, 2003; Zakaria & Nordin, 2008). When the research on mathematics anxiety (Brady & Bowd, 2005; Brown, Westenskow, & Moyer-Packenham, 2011; Gresham, 2007; Jackson, 2008; Liu, 2008; Tooke & Lindstrom, 1998; Uusimaki & Nason, 2004; Vinson, 2001) is reviewed, it is observed that the majority of students and pre-service teachers have mathematics anxiety, and various mathematical materials and manipulatives (Vinson, 2001) are used and mathematics teaching courses (Gresham, 2007) are organized in order to reduce this anxiety level.
When we examine the literature about pre-service teachers’ mathematics anxiety, we observe that it generally signifies weak mathematical backgrounds, histories and experiences; it seems instinctual that teachers who possess negative feelings and abilities in any subject domain, such as mathematics, would have trouble in teaching the subject to students. Mathematics anxiety can be regarded as a pre-existing condition or a negative mathematics trait that pre-service teachers experience when they begin teacher training programs. Nevertheless, this emphasis on mathematics anxiety as a pre-existing condition disregards the anxiety that may emerge due to teaching mathematics or mathematics teaching anxiety (Brown, Westenskow, & Moyer-Packenham, 2011).

While mathematics anxiety is among the fears that students have, it is considered that pre-service teachers may have both mathematics anxiety and encounter teaching anxiety as they approach the end of their undergraduate period. Pre-service teachers’ teaching anxiety is of such great significance in mathematics courses that they experience difficulty in learning and teaching (Peker, 2006). Mathematics teaching anxiety can be described as pre-service and in-service teachers’ moods of tension and anxiety emerging in teaching mathematical concepts, theories and formulas or problem solving (Peker, 2009). Mathematics teaching anxiety can be unrelated to an individual’s insufficient mathematics history or background. Thus, a person may not encounter mathematics anxiety and may be very self-reliant about their mathematics knowledge, but they may encounter mathematics teaching anxiety because they are not self-reliant about their ability to teach children the mathematics that they know (Brown, Westenskow, & Moyer Packenham, 2011). The symptoms of mathematics teaching anxiety may contain intense nervousness, the inability to concentrate, negative self-talk, being easily distressed by noises, being unable to hear the students, and sweaty palms (Peker, 2009). Elmas (2010) found that mathematics anxiety, traineeship, lack of self-confidence and content knowledge were the factors that caused mathematics teaching anxiety in pre-service teachers. Peker and Halat (2009) researched the effects of mathematical visualization activities formed with WebQuest and spreadsheets on pre-service elementary teachers’ mathematics teaching anxiety. From this research, it was observed that WebQuest activities were more effective in reducing pre-service teachers’ teaching anxiety. It is observed that these activities that affect teaching anxiety levels may be reduced thanks to technological facilities. It was observed that technology use in mathematics teaching (Berger, 2010; Erbaş & Aydoğan Yenmez, 2011; Güven & Karataş, 2003; Lagrange, 1999; Marshall, Buteau, Jarvis, & Lavieza, 2012; Wong, Yin, Yang, & Cheng, 2011) positively contributed to anxiety levels experienced by students and teachers. As a result of technological advances, one contribution is the active role played by Webquest activities in reducing mathematics teaching anxiety (Peker & Halat, 2009).

Ersoy (2005) stated that a teacher’s consciousness level and beliefs have an important effect on not only starting but also stopping or preventing a reform movement in schools. For this reason, he emphasized that how teachers perceive and evaluate cognitive tools and new educational technologies from their point of view is among the important issues that must not be overlooked in a reform movement. That is to say, how mathematics teachers of the future perceive this technology is an important issue that we face.

Computer literacy and internet use have become a necessity for both educators and students in order for technological developments to be effectively used in education. On the other hand, the use of computer technologies has become inevitable for every field of occupation in developed societies, and computer literacy has emerged as an important skill for individuals in modern society (Yanık, 2010). He stated that it is important to provide coherent training in computer literacy beginning from the last year of elementary education and continuing with secondary education, and previously acquired computer knowledge and skills must be developed in university years for the purpose of research and problem solving (Yazıcı, 2001).

Nowadays, the correct and effective use of technology in the teaching process by teachers contributes to participating in courses in a productive manner. Application of technology develops a concrete and experimental approach to mathematics subjects, and enables students to develop more abstract and symbolic skills (Flores, 2002). There are various researches on the impact of technology on success (Tajuddin, Tarmizi, Konting & Ali, 2009) anxiety (Uusimaki & Nason (2004), motivation (Nordin, Zakaria, Mohamed & Embi, 2010) and mathematics learning and teaching activities (Fahlberg-Stojanovska & Trifunov, 2010).

In-depth description and analysis of the influence of technology on mathematics teaching anxiety will be a significant step. Identifying the main factors that the impact of technology has on mathematics teaching anxiety is the focus of the research. How technology will affect pre-service teachers’ teaching anxiety is important because it offers practical information on teacher training. It is easier and more meaningful to provide the required skills to pre-service teachers during their university education rather than after they start teaching. This way, we can gather
information about the mathematics teaching anxiety levels of pre-service teachers and define variables related to teaching anxiety. In view of all these factors, the nature of the relationship between technology and mathematics teaching anxiety has caught the attention of researchers. In this context, the aim of this research is to determine the relationship between pre-service teachers’ mathematics teaching anxiety and their perceptions of technology use in mathematics teaching as well as their computer literacy levels.

**Method**

The nonexperimental correlational research, which is included in the quantitative research approach, was used in the study. Correlational research involves studies conducted to determine relationships between two or more variables and to obtain clues about cause and effect (McMillan & Schumacher, 2010).

**Sample**

This research was conducted with 481 elementary and secondary pre-service mathematics teachers who were selected using the random sampling method. The sample of the research composed of 320 pre-service teachers who were studying in four classes in the elementary mathematics teaching program and 161 pre-service teachers who were studying in five classes in the secondary mathematics teaching program. All the pre-service teachers volunteered to participate in the research. Table 1 shows the frequency of 463 pre-service teachers who completely filled all scales out of the 481 pre-service teachers who participated in the study.

| Table 1. Frequency of Class Levels and Teaching Area Distribution of Pre-Service Teachers |
|-----------------------------------------------|-----------------------------------------------|
| Class Level | Elementary | Teaching Area |            | Secondary |            |
|             | f           |              |            | f           |              |
| First Class | 97 (M:43-F:54) | 34(M:15-F:19) |            |            |              |
| Second Class | 65(M:19-F:46) | 29(M:9-F:20) |            |            |              |
| Third Class | 76(M:30-F:46) | 40(M:17-F:23) |            |            |              |
| Fourth Class | 75(M:36-F:39) | 26(M:12-F:14) |            |            |              |
| Fifth Class | — | 21(M:12-F:9) |            |            |              |
| Total       | 313 (M:128-F:185) | 150 (M:66-F:84) |            |            |              |

**Data collection tools**

In the research, three instruments were used namely, (1) the Mathematics Teaching Anxiety Scale (MATAS) to determine pre-service teachers’ levels of mathematics teaching anxiety, (2) the Perception Scale for Technology Use in Mathematics Teaching (PSTM) to determine their perceptions regarding technology use in mathematics courses and (3) the Computer Literacy Scale (CLS) to determine their computer literacy levels. Two factors were taken into consideration in selecting these three scales: sufficiency of validity and reliability studies and the latest scales in Turkish language.

MATAS, which was developed by Peker (2006), consists of 23 items in four sub-dimensions with a 5-point Likert-type scale ranging from “1” for “completely disagree” to “5” for “completely agree.” These sub-dimensions are as follows: anxiety arising from content knowledge (10 items), anxiety arising from self-confidence (6 items), anxiety arising from attitude towards mathematics teaching (4 items) and anxiety arising from teaching knowledge (3 items). There are 10 negative items in MATAS and these items are scored by reverse calculation. Peker (2006) found that the reliability coefficient (Cronbach’s alpha) of the scale was .91. The internal consistency coefficients for each dimension were .90 for the dimension of anxiety arising from content knowledge, .83 for the dimension of anxiety arising from self-confidence,.71 for the dimension of anxiety arising from attitude towards mathematics teaching and .61 for the dimension of anxiety arising from teaching knowledge. In this study, the reliability coefficient obtained for the complete instrument was .92. Regarding each dimension, the reliabilities were found to be .87 for anxiety arising from content knowledge, .84 for anxiety arising from self-confidence, .88 for anxiety arising from
attitude towards mathematics teaching and .80 for anxiety arising from teaching knowledge. In the scale, the total score and the scores achieved in each dimension were also computed. The highest score that could be achieved on the Mathematics Teaching Anxiety Scale is 115, whereas the lowest score is 23 and the average score is 69. The highest score in the content knowledge sub-dimension is 50, the lowest score is 10, and the average score is 30. The highest score in the self-confidence sub-dimension is 30, the lowest score is 6, and the average score is 18. The highest score in the attitude towards mathematics teaching sub-dimension is 20, the lowest score is 4, and the average score is 12. The highest score in the teaching knowledge sub-dimension is 15, the lowest score is 3, and the average score is 9. The high scores obtained in this study indicate that the pre-service teachers’ mathematics teaching anxiety was high. Table 2 shows the number of items in MATAS sub-dimensions, sample items for each dimension and reliability coefficients.

Table 2. The details of sub-dimensions of MATAS

<table>
<thead>
<tr>
<th>Sub-dimensions of MATAS</th>
<th>Number of items</th>
<th>Sample items</th>
<th>Reliability coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content knowledge</td>
<td>10</td>
<td>In my teaching profession, I think I feel desperate while teaching mathematics subjects</td>
<td>.87</td>
</tr>
<tr>
<td>Self-confidence</td>
<td>6</td>
<td>I feel competent about solving mathematics problems while teaching.</td>
<td>.84</td>
</tr>
<tr>
<td>Attitude towards mathematics teaching</td>
<td>4</td>
<td>I think teaching mathematics subjects is fun for me.</td>
<td>.88</td>
</tr>
<tr>
<td>Teaching knowledge</td>
<td>3</td>
<td>I can make use of special teaching strategy knowledge and skills while teaching mathematics.</td>
<td>.80</td>
</tr>
</tbody>
</table>

PSTM, which was developed by Öksüz, Ak and Uça (2009), consists of 73 items in three sub-dimensions with a 5-point Likert-type scale ranging from “1” for “completely disagree” to “5” for “completely agree.” The three sub-dimensions are requirement, advantages and disadvantages. The internal consistency coefficient of the scale was .96 (Öksüz, Ak, & Uça 2009). The internal consistency coefficients obtained for the sub-dimensions were .95 for requirement, .96 for advantages; and .84 for disadvantages. In this research, the internal consistency coefficient of the scale was found to be .94 while the internal consistency coefficients obtained for each of the sub-dimensions were .88, .93 and .61 respectively. The highest score that can be achieved in the scale is 365 whereas the lowest is 73. Items in the disadvantages sub-scale are done by reverse calculation. The total score that can be achieved by reverse grading of negative items and high score for each dimension define positive perception, while low score defines negative perception. Table 3 shows the number of items in PSTM sub-dimensions, sample items for each dimension and reliability coefficients.

Table 3. The details of sub-dimensions of PSTM

<table>
<thead>
<tr>
<th>Sub-dimensions of PSTM</th>
<th>Number of items</th>
<th>Sample items</th>
<th>Reliability coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Requirement</td>
<td>29</td>
<td>The Following Software Are Required for Mathematics Teaching; “specific mathematics software (Sketchpad, Cabri etc.)”</td>
<td>.88</td>
</tr>
<tr>
<td>Advantages</td>
<td>34</td>
<td>Use of technology in mathematics teaching facilitates teaching with a constructivist approach</td>
<td>.93</td>
</tr>
<tr>
<td>Disadvantages</td>
<td>10</td>
<td>Use of technology in mathematics teaching causes a waste of time.</td>
<td>.61</td>
</tr>
</tbody>
</table>

CLS, which was developed by Kay (1990) and translated into Turkish by Kılınç and Salman (2006), was used to determine the pre-service mathematics teachers’ computer literacy levels. The Computer Literacy Scale consists of 24 items with a 7-point Likert-type scale ranging from “1” for “strongly disagree” to “7” for “strongly agree.” The scale is composed of four dimensions, namely basic skills, application software skill, programming and computer awareness. Kılınç and Salman (2006) found that the internal consistency coefficients obtained for the sub-dimensions were .91 for basic skills, .93 for application software skills .91 for programming and .94 for computer awareness. Based on the data of this research, the internal consistency coefficient of the scale was found to be .92 while the internal consistency coefficients for each of the dimensions were .89, .88, .88 and .81, respectively. The highest score that can be achieved is 168 whereas the lowest score is 24. In this study the above average scores signify that the pre-
service teachers’ computer literacy levels were positive. Table 4 shows the number of items in the CLS sub-dimensions, sample items for each dimension and reliability coefficients.

<table>
<thead>
<tr>
<th>Sub-dimensions of CLS</th>
<th>Number of items</th>
<th>Sample items</th>
<th>Reliability coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic skills</td>
<td>6</td>
<td>I can format a new diskette</td>
<td>.89</td>
</tr>
<tr>
<td>Application software</td>
<td>6</td>
<td>I can teach someone to use a computer software package</td>
<td>.88</td>
</tr>
<tr>
<td>Programming</td>
<td>6</td>
<td>I can write a computer program in BASIC or Logo</td>
<td>.88</td>
</tr>
<tr>
<td>Computer awareness</td>
<td>6</td>
<td>I can identify the basic parts of computers and their functions</td>
<td>.81</td>
</tr>
</tbody>
</table>

**Data analysis**

The analysis was performed for 463 pre-service teachers who completed all three instruments. The SPSS 16.0 statistics package program was used for data analysis. Arithmetic mean and standard deviation values were used in descriptive statistics. Normality analysis was initially conducted on the data in order to determine the most appropriate test that could be used in analyzing the data. While conducting the analysis, the Kolmogorov-Smirnov test was used since the number of samples for each measurement was higher than 29 (Kalaycı, 2010). It was found that the data were not distributed normally in the Perception Scale for Technology Use in Mathematics Teaching, Mathematics Teaching Anxiety Scale, Computer Literacy Scale and the sub-dimensions of these scales ($p < .05$). Therefore, Spearman’s Rank Correlation Coefficient was used when examining the relationships.

**Findings**

The scores of pre-service mathematics teachers in MATAS including its sub-dimensions are given in Table 5.

<table>
<thead>
<tr>
<th>Sub-dimensions of MATAS</th>
<th>n</th>
<th>$\bar{X}$</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content knowledge</td>
<td>463</td>
<td>18.63</td>
<td>5.61</td>
</tr>
<tr>
<td>Self-confidence</td>
<td>463</td>
<td>13.18</td>
<td>3.61</td>
</tr>
<tr>
<td>Attitude towards teaching</td>
<td>463</td>
<td>6.93</td>
<td>2.31</td>
</tr>
<tr>
<td>Teaching knowledge</td>
<td>463</td>
<td>5.88</td>
<td>1.90</td>
</tr>
<tr>
<td>Total MATAS</td>
<td>463</td>
<td>44.64</td>
<td>.50</td>
</tr>
</tbody>
</table>

Given the overall average score of pre-service teachers in the mathematics teaching anxiety scale ($\bar{X} = 44.64$), we can conclude that they have low levels of anxiety. Considering the content knowledge ($\bar{X} = 18.63$), self-confidence ($\bar{X} = 13.18$), attitude ($\bar{X} = 6.93$) and teaching knowledge ($\bar{X} = 5.88$) average scores of pre-service teachers given in Table 5, we can conclude that they have low levels of anxiety overall in all sub-dimensions.

In this section the relationships between pre-service teachers’ mathematics teaching anxiety and their perceptions regarding technology use in mathematics teaching as well as their computer literacy levels were analyzed. Spearman’s Rank Correlation Coefficients of the scores achieved in mathematics teaching anxiety and perceptions regarding technology use in mathematics teaching are given in Table 6.

As seen in Table 6, a low-level, negative and significant relationship was found between mathematics teaching anxiety and pre-service teachers’ perceptions regarding technology use in mathematics teaching ($r = .240; p < .01$). Furthermore, a low-level, negative and significant relationship was found between the mathematics teaching anxiety and “requirement” ($r = .247; p < .01$) and “advantages” ($r = .156; p < .01$) sub-dimensions of pre-service teachers’ perceptions regarding technology use. On the other hand, a low-level, positive and significant relationship was found between the mathematics teaching anxiety and disadvantages sub-dimensions of pre-service teachers’ perceptions
regarding technology use \((r = .150; p < .01)\). In view of this finding, it can be stated that pre-service teachers who had positive perceptions regarding technology use in mathematics teaching had lower mathematics teaching anxiety. When the sub-dimensions were examined, it was observed that pre-service teachers who had positive perceptions in the “requirement” and “advantages” sub-dimensions of technology use had lower teaching anxiety. Similarly, it can be stated that pre-service teachers’ teaching anxiety increased as their perception of disadvantages regarding technology use increased. A low-level, negative and significant relationship was found between perceptions regarding technology use in all sub-dimensions of mathematics teaching anxiety. The highest-level relationship between variables was found between “teaching knowledge” that is the sub-dimension of mathematics teaching anxiety and perceptions regarding technology use in mathematics teaching. This relationship is negative and significant \((r = -.352; p < .01)\). In view of this, it can be stated that pre-service teachers who had positive perceptions regarding technology use in mathematics teaching were less anxious in the dimension of teaching knowledge. When the relationship between other sub-dimensions in Table 1 was examined, no significant relationship of mathematics teaching anxiety was found between pre-service teachers’ self-confidence and their scores in perceptions regarding advantages of technology use in mathematics teaching \((r = .078; p > .05)\) as well as their scores in disadvantages of technology use in mathematics teaching \((r = .077; p > .05)\). Similarly, no significant relationship on mathematics teaching anxiety was found between pre-service teachers’ attitudes toward mathematics teaching and their scores in perceptions regarding disadvantages of technology use in mathematics teaching \((r = .065; p > .05)\). Apart from these, a negative and significant relationship was found between other variable pairs.

**Table 6.** Relationship between mathematics teaching anxiety and perceptions regarding technology use in mathematics teaching \((N = 463)\)

<table>
<thead>
<tr>
<th>Sub-dimensions of MATAS</th>
<th>Variable</th>
<th>PST</th>
<th>Sub-dimensions of PSTM</th>
<th>Requirement</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MATAS</td>
<td>-.240**</td>
<td>- .247**</td>
<td>-.156**</td>
<td>.150**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Content knowledge</td>
<td>-.171**</td>
<td>-.184**</td>
<td>-.096**</td>
<td>.149**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Self-confidence</td>
<td>-.135**</td>
<td>-.139**</td>
<td>-.078</td>
<td>.077</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Attitude towards</td>
<td>-.247**</td>
<td>-.274**</td>
<td>-.185**</td>
<td>.065</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mathematics teaching</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teaching knowledge</td>
<td>-.352**</td>
<td>-.316**</td>
<td>-.307**</td>
<td>.141**</td>
<td></td>
</tr>
</tbody>
</table>

\(p < .05. \quad *p < .01.\)

Spearman’s Rank Correlation Coefficients of scores in mathematics teaching anxiety and computer literacy are given in Table 7.

**Table 7.** Relationship between teaching anxiety and computer literacy \((N = 463)\)

<table>
<thead>
<tr>
<th>Sub-dimensions of MATAS</th>
<th>Variable</th>
<th>CLS</th>
<th>Basic skills</th>
<th>Application software skills</th>
<th>Programming</th>
<th>Computer awareness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MATAS</td>
<td>-.247**</td>
<td>-.260**</td>
<td>-.280**</td>
<td>-.010</td>
<td>-.264**</td>
</tr>
<tr>
<td></td>
<td>Content knowledge</td>
<td>-.192**</td>
<td>-.195**</td>
<td>-.227**</td>
<td>-.006</td>
<td>-.218**</td>
</tr>
<tr>
<td></td>
<td>Self-confidence</td>
<td>-.259**</td>
<td>-.260**</td>
<td>-.270**</td>
<td>-.048</td>
<td>-.235**</td>
</tr>
<tr>
<td></td>
<td>Attitude towards</td>
<td>-.166**</td>
<td>-.176**</td>
<td>-.226**</td>
<td>-.014</td>
<td>-.200**</td>
</tr>
<tr>
<td></td>
<td>mathematics teaching</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teaching knowledge</td>
<td>-.224**</td>
<td>-.245**</td>
<td>-.229**</td>
<td>-.023</td>
<td>-.234**</td>
</tr>
</tbody>
</table>

\(p < .05. \quad *p < .01.\)

A low-level, negative and significant relationship \((r = .247; p < .01)\) was found between mathematics teaching anxiety and pre-service mathematics teachers’ computer literacy (see Table 7). In view of this, it can be stated that pre-service mathematics teachers’ mathematics teaching anxiety decreased as their computer literacy increased. A low-level, negative and significant relationship was found between pre-service teachers’ mathematics teaching anxiety, all sub-dimensions of this anxiety and other dimensions excluding the programming sub-dimension of computer literacy. The highest-level relationship was found between pre-service teachers’ general mathematics teaching anxiety and the application software skills sub-dimension of computer literacy. This relationship is negative and significant \((r = .280; p < .01)\). No significant relationship was found between the pre-service mathematics teachers’ programming sub-dimension of computer literacy and their general mathematics teaching anxiety \((r = .010;\)
terms of content knowledge, self-confidence, attitudes towards mathematics teaching, and teaching knowledge sub-dimensions of mathematics teaching anxiety levels of pre-service teachers were described. Accordingly, it was found that pre-service teachers had low mathematics teaching anxiety levels in each sub-dimension. It was found that pre-service teachers who had positive perceptions regarding technology use had lower teaching anxiety shows parallelism with the result of the study conducted Sun and Pyzdrowski (2009). Also, Taylor and Galligan (2006) have found that using technology in mathematics classroom results in students displaying positive attitude toward mathematics.

It was found that levels of teaching anxiety in terms of content knowledge, self-confidence, attitude towards mathematics teaching and teaching knowledge could decrease among pre-service mathematics teachers who had positive perceptions regarding technology use. For this reason, in order to reduce mathematics teaching anxiety and increase technology awareness, necessary attention must be paid to the integration of technology into mathematics teaching and learning in the courses that the pre-service teachers take throughout their university education. Apart from this, it is considered that courses on how to use technology in mathematics teaching will positively contribute to the pre-service teachers’ perceptions regarding technology, thereby reducing their mathematics teaching anxiety.

It was found that positive perceptions in the “requirement” and “advantages” sub-dimensions of technology use reduced teaching anxiety, and teaching anxiety increased as the perception in the “disadvantages” sub-dimension increased. It could be deduced from this fact that those teachers who think technology is necessary and advantageous in teaching and learning have lower teaching anxiety whereas those who think technology is disadvantageous have higher teaching anxiety. Therefore, future qualitative studies conducted with those teachers who think that technology use is disadvantageous in mathematics teaching and learning should identify the reasons underlying this condition. Information from these studies will provide more evidence for efforts to reduce teaching anxiety among teachers.

This study also examined the relationship between pre-service teachers’ computer literacy levels and mathematics teaching anxiety. It was found that mathematics teaching anxiety decreased as computer literacy levels increased. Sun and Pyzdrowski (2009) asserted that anxiety decreases in an environment in which students can build their own knowledge and learn mathematics cooperatively using computers, various software and websites with mathematical content. Considering that mathematics anxiety and mathematics teaching anxiety are related (Peker & Ertekin, 2011), the results obtained in the research that mathematics teaching anxiety could decrease as computer literacy levels increase shows parallelism with the result of the study conducted Sun and Pyzdrowski (2009).

It was found that pre-service mathematics teachers’ teaching anxiety in terms of content knowledge, self-confidence, attitude towards mathematics teaching and teaching knowledge decreased as their computer literacy levels increased. Furthermore, it was found that pre-service teachers’ mathematics teaching anxiety could decrease as their computer literacy in the sub-dimensions of basic skills, application software skills and computer awareness increases. In view of these findings, there is an urgent need to develop pre-service teachers’ computer literacy in order to reduce their mathematics teaching anxiety.
Considering that pre-service teachers’ mathematics teaching anxiety may decrease as their perception levels regarding technology use in mathematics teaching positively increase and their computer literacy levels increase, it is considered that prioritizing the applications in which technology is used in the teacher training process, and supporting these applications with up-to-date software, will contribute significantly to decreasing teaching anxiety. In the literature, there are various studies on teaching anxiety in English, Music and Psychology fields (Gardner & Leak, 1994; Williams, 1991; Strong, 2013). For instance, training programs such as a consultant observation program and peer mentoring activities in English are found to contribute to decreasing teaching anxiety (Williams, 1991). Similarly, the importance of teacher training programs is also emphasized in psychology (Gardner & Leak, 1994). Therefore, training programs employing latest technologies can play a critical role in reducing mathematics teaching anxiety.

References


Validating the Learning Cycle Models of Business Simulation Games via Student Perceived Gains in Skills and Knowledge

Yu-Hui Tao\(^1\), C. Rosa Yeh\(^2\) and Kung Chin Hung\(^3\)

\(^1\)Department of Information Management, National University of Kaohsiung, 700 Kaohsiung University Road, Nan-Tzu District, Kaohsiung, Taiwan / \(^2\)Graduate Institute of International Human Resource Development, National Taiwan Normal University, 162, Sec. 1, Heping E. Rd., Taipei, Taiwan / \(^3\)Top-BOSS International Corp., 7F., No.153, Sec. 3, Xinyi Rd., Taipei, Taiwan; Department of International Trade, Chihlee Institute of Technology, Taiwan // ytao@nuk.edu.tw // rosayeh@ntnu.edu.tw // kc@top-boss.com.tw

\(^*\)Corresponding author

(Submitted February 11, 2014; Revised June 7, 2014; Accepted July 17, 2014)

ABSTRACT
Several theoretical models have been constructed to determine the effects of business simulation games (BSGs) on learning performance. Although these models agree on the concept of learning-cycle effect, no empirical evidence supports the claim that the use of learning cycle activities with BSGs produces an effect on incremental gains in knowledge and skills. Therefore, this study aims to provide quantitative evidence by conducting an experiment using BSGs with different complexity levels in an undergraduate general course. Three research propositions guide a corresponding experimental design to collect data on student perception. Statistical analyses of the perception data from 43 student respondents reveal that skills, declared knowledge, procedural knowledge, and strategic knowledge greatly increase in the initial cycles but vary in the rate of increase at later cycles. This research, which adds to the literature of game learning models, concludes that learning cycles with different activities are required to sustain the development of knowledge and skills acquired through BSGs. The empirical outcome also serves as a useful reference for teachers who are planning to adopt BSGs in their class activities.

Keywords
Business simulation game, Game model, Learning cycle, Knowledge gain, Skill gain

Introduction

Background

Business simulation games (BSGs) are widely available since 1961 (Kibbee, Craft, & Nanus, 1961). For more than a decade, its usage has reached 97.5% among all Association to Advance Collegiate Schools of Business member schools in the United States; furthermore, BSGs have obtained high adoption rates in universities in countries such as United Kingdom and Australia (Faria & Nulsen, 1996). Prensky (2003) claimed that games provide an indispensable motivational condition for student learning and denoted that students are more interested in choosing courses that incorporate BSGs. Therefore, BSG has been promoted as a popular means of learning in both informal (Kapp, 2006) and formal settings (Lim, 2008; Prensky, 2008).

As summarized in Tao, Yeh, and Hung (2012), a simulation game combines the features of both a game and a simulation. In general, the goal of a game is to determine win or lose, and that of a simulation is to identify the relationship between causes and effects of a decision. Specifically, a game is a competition that involves competitors, constraints, a linear objective and playability, whereas a simulation represents a real causal relationship and may have a non-linear objective and non-stop status. BSG may be referred as a simulation game for business learning by playing. Most BSGs require learners to form teams representing different functional groups and make group decisions (Feinstein, Mann, & Corsun, 2002). Traditional BSGs are usually designed to be played in rounds (Tao, Yeh, & Hung, 2012), and decisions for the next round are modified based on the evaluation of the outcomes from decisions made during the previous round. After a certain number of rounds, the ranking of all the competing teams is determined according to the pre-defined performance criteria.

Farrell (2005) illustrated a learning environment which incorporates simulations in an international business course, as shown in Figure 1. In this environment, simulations with interactive, participative, inductive, reflective, and
exploratory characteristics can be adopted with traditional pedagogical methods as supplements, including classroom lectures, textbooks, case studies, and research papers, to stimulate student interest and involvement and to accomplish the desired learning objectives.

Figure 1. Learning environment for a course using simulation by Farrell (2005)

Early studies on the relationship between student perception or attitude of BSG learning and conventional learning performance (e.g., course grade) present mixed results (Garris, Ahlers, & Driskell, 2002; Tao, Yeh, & Hung, 2012). Scholars such as Anderson and Lawton (2007) thus suggested finding the missing links between simulation performance and student attitude. However, according to the research findings of Tao, Yeh, and Hung (2012), learning performance by different assessment methods such as tests, competition, and essay sharing in a BSG class, is influenced more by the different levels of student characteristics in motivation and persistence, learning preference, and learning attitude. Therefore, it is understandable that recent studies on BSG focus more on non-performance outcomes, such as student attention and class engagement.

Research rationale and objectives

Several theoretical models have already been proposed for game learning. Among these, two game learning models, the problem-based game (PBG) model by Kiili (2007) and the input–process–outcome (IPO) model by Garris et al. (2002), utilized multiple learning cycles to increase the learning effects of BSGs. The following research issues can be observed regarding the above-mentioned game models. First, the research methods of these models are both qualitative-oriented; that is, no quantitative analyses are given to support their models. Second, only one BSG (i.e., Realgam) is involved in the study by Kiili (2007), which may not be representative of other BSGs. The training simulation application used in the study by Garris et al. (2002) is not a BSG at all. Third, extant BSG literature focuses on business major students, however, debates exist on the relationship between business knowledge and BSG performance (Goosen, 2002; Wolfe & Luethge, 2003).

This research aims to conduct a quantitative experiment to validate the effect of the learning cycles (Issue 1) of three different complexity levels of BSGs (Issue 2) on the basis of the perceptions of college students with different majors in Taiwan (Issue 3). In particular, the learning cycle is investigated by arranging a series of learning activities to trigger certain game strategy changes in the double-loop learning suggested by Kiili (2007) and in the user judgment in the game cycle proposed by Garris et al. (2002). The present research focuses on student perception and disregards common learning performance indicators because of two reasons: First, there are mixed results regarding the relationship between learning performance and student perception in literature and in practice. Second, it seems that BSG learning performance is more fully presented in literature, while empirical tests of student perception or attitude toward BSG are somewhat scarce. That’s why recent research and modern classroom management focus more on the aspect of student perception. Related literature is presented to derive appropriate research methods to address the three research issues. The corresponding results, discussions, and conclusions are also given in the remaining parts of the paper.
Literature review

Following the research rationale and objectives above, we will briefly review the literature in BSG related theoretical models, BSG complexity, and BSG in Taiwan as the basis for the research design.

Theoretical models for playing BSGs

As shown in Figure 2, Kiili (2007) proposed a problem-based gaming (PBG) model, in which learning in PBG is defined as the construction of cognitive structures through actions in the game world. Specifically, the PBG model describes learning as a cyclic process through direct experience in the game world. PBG distinguishes the learning process into elements of strategy formation, active experimentation, game world observation, and reflection. In the beginning of the game, the player forms a playing strategy according to the player’s prior experiences. During this process, single-loop learning is formed if the player goes from reflection to active experimentation without forming new strategies. By contrast, double-loop learning is formed when reflections are followed by forming new strategies. Feedback is given regarding the action of the player to support reflective thinking and knowledge construction by focusing the player’s attention to relevant information.

![Figure 2. Problem-based gaming model by Kiili (2007)](image)

According to Kiili (2007), learning outcome is a critical component in the game model because the outcome of the reflection “can be personal synthesis or appropriation of knowledge, validation of hypothesis laid during playing strategy formation or a new strategy to be tested.” Several broad categories of learning outcomes were classified by Kraiger, Ford, and Salas (1993), including skill-based, cognitive, and affective outcomes. As seen in Figure 3, skill-based learning outcomes include compilation and automaticity, which address technical or motor skills. Cognitive learning outcomes include verbal knowledge, knowledge organization, and cognitive strategies, which represent three subcategories of declarative knowledge, procedural knowledge, and strategic knowledge. Affective learning outcomes refer to attitudes that include attitudinal and motivational aspects.

Input-process-outcome (IPO) is another game learning model proposed by Garris et al. (2002). As seen in Figure 4, both the iterative game cycles presented by Kiili (2007) and the three-outcome categories mentioned by Kraiger et al. (1993) are included in the IPO game model. Specifically, in this game model, instructional content and game characteristics are paired in the input design of repeated judgment–behavior–feedback game cycles to achieve skill-based, cognitive, and affective learning outcomes. Moreover, the learner constructs knowledge from experience in the IPO.

Garris et al. (2002) summarized a user self-directed tacit model of learning on the basis of studies on instructional games. This model aims to design an instructional game that incorporates several game features or characteristics. These features trigger a cycle that includes user judgments or reactions (e.g., enjoyment or interest), user behavior (e.g., greater persistence or time on task), and further system feedback. The model of Garris et al. pairs instructional content with appropriate game features; thus, this game cycle results in a recurring and self-motivated game play. Engagement in game play achieves the intended training objectives and specific learning outcomes.
Complexity of BSGs

The research issue of BSG complexity received much attention after Cannon (1995) declared the complexity paradox. In recent literature, complexity is still a research issue or element; for instance, Lainema and Lainema (2007) identified complexity as one of the five elements that enhance knowledge acquisition through simulation environment.

Several studies have diversified the complexity issue by promoting the concept of learning design or by constructing games for better learning effects. Prensky (2008) and Lim (2008) recommended learners to design games to achieve closely their learning needs and desired performance. In another study, utilizing game complexity as a variable, Vos and Brennan (2010) validated that game construction could effectively enhance student motivation and in-depth learning.

Few other studies have directly addressed the complexity issue in BSG learning. Tech and Murff (2008) concluded that using complex and real teaching simulation games may not be advantageous and that the learning effects brought by small BSGs may be more significant than that of big BSGs. Accordingly, several recent studies have attempted to develop a multi-module BSG software to adjust the complexity level of BSGs through combination or repetition of modules. Chang, Chen, Yang, and Chao (2009) developed a production and logistic management game with flexibly
assembled modules. In another study, Yasarcan (2010) suggested a gradual-increase-in-complexity approach to improve the performance, learning, and understanding level of dynamic managing BSGs. Tao, Yeh, and Hung (2012) considered game complexity as a parameter in their experiment when exploring the learning performance of BSGs.

BSGs in Taiwan

Imported and domestic BSGs are both available in the Taiwanese market, both of which are provided by two major BSG providers, namely, Top-Boss Corporation (http://www.top-boss.com.tw) and Pitotech Corporation (http://www.pitotech.com.tw). According to the review of BSG development by Faria and Nulsen (1996), universities in the United Kingdom develop in-house BSG, whereas universities in the United States use mainly published simulations. Similarly, Taiwan Pitotech mainly sells imported simulations, whereas Top-Boss develops many in-house games. For example, BOSS is a total-business game that involves all functional areas in business organizations to develop the business decision-making skills of students; thus, BOSS is appropriate in illustrating the knowledge and applications of targeted course subjects in a classroom setting (Prensky, 2008).

Smaller BSGs, such as Web-based games, are recently available and are becoming popular in Taiwan. Retail Expert is a convenient store management game that was first developed by Top-Boss. Given the high acceptance and popularity of Retail Expert, Top-Boss quickly produced many other small BSGs, such as IPTTour, a financial management game, and COO-legend, a restaurant management game promoted by a national competition activity upon initial market launch (http://www.top-boss.com.tw/joomla/index.php?option=com_content&task=view&id=473&Itemid=1).

Research design

This section will first state the research propositions, and then describe the experiment and the questionnaire design. Finally, the statistical methods used in this study will be presented.

Research propositions

Three research propositions are formulated on the basis of the research objective and literature review:
Proposition 1: The perceived learning cycle effects are equivalent across different complexity levels of BSGs in a student sample with diverse background.
Proposition 2: The perceived incremental learning cycle effects are significant across different complexity levels of BSGs in a student sample with diverse background.
Proposition 3: The perceived leaning cycle effects match with the competition outcomes across different complexity levels of BSGs in a student sample with diverse background.

Experiment design

To test the research propositions, an experiment was designed to collect data from college students with different majors to validate the perceived learning-cycle effects caused by playing multiple BSGs with different complexity
levels. Consequently, a general course called Business Simulation Games and Competition was offered to all undergraduate students in a national university in southern Taiwan. This course enabled students to play three BSGs with different complexity levels (Tech & Murff, 2008), namely, BOSS (total enterprise simulation), Retail Expert (single-player small game), and Beer Game (multiple-player small game). Class activities were designed to revolve around repeated game cycles to trigger the learning effects of the PBG model by Kiili (2007) and the IPO game model by Garris et al. (2002). Both models involve repeated cycles such that certain activities can be incorporated into the learning-by-playing process.

Many combinations of learning activities can be used in the collection of learning-cycle effects within this experimental setting. This study considers a few commonly recognized learning activities which each game follows: (i) game introduction, (ii) game practice via simulated competition, (iii) group thought-sharing report, and (iv) final game competition. The participating students have different majors; thus, students may have or may not have a background in management. The learning effects within this learning-by-playing scenario provide opportunities for peer collaboration, exploration of useful information, and competitive pressure among students. Many game tips can also be learned from winners who share their winning experience after each practice and the formal competition, as well as from the thought-sharing reports that provide necessary knowledge, tactics, and strategies.

A self-report questionnaire is administered at the end of the class to collect the perception of student participants regarding the BSGs. The questionnaire design is described below.

**Questionnaire design**

Because of the nature of the research objectives on validating the BSG process models, the questionnaire design is an after-experiment survey on student perceptions, which is consisted of two parts: student profile and measurement items on student perceptions. The first part, which focuses on intended student profile, includes gender, management major, previous management courses, previous BSG experience, and average scores from the last semester. Gender has been associated with learning outcomes and has been investigated in the BSG context in previous studies (Hawk & Shah, 2007). Given that previous business experience plays an important role in BSG performance (Hawk & Shah, 2007), previous management courses and previous BSG experience are used in this study to reflect the business experience of the participants. The sample students who have participated in this study came from five different disciplines, including Management, Science, Engineering, Society and Humanities, and Law. The options for each profile question are presented in detail in the later section of Data Analysis.

The second part of the questionnaire, which measures student perceptions, consists of student-perceived complexity level of BSGs, gains on skills and each type of knowledge after each learning activities, and matching of these gains with the competition outcome. Two straightforward items by Tao, Yeh, and Hung (2012) that measure the relative BSG complexity and simplicity are used. The two items, namely, “I think the game is complicated” and “I think the game is easy to use,” are included for each of the three BSGs in this course. Following the study of Kraiger et al. (1993), the present research addresses the cognitive and skill-based outcomes, in which knowledge is categorized into declarative, procedural, and strategic knowledge, whereas skills are simplified as general skills. For each category under declarative knowledge, procedural knowledge, strategic knowledge, and skills, the perceptions of the student participants are collected to assess their incremental gains at five different activity stages, namely, before introducing the game, after introducing the game, after informal game practicing and sharing, after the group has written the thought-sharing report, and after the final competition. The example questionnaire items are similar to “My game skills improved after informal game practicing and sharing.” The perceptions of participants regarding how the BSG competition outcomes match their gains in knowledge and skills are also collected. An example questionnaire item is “In general, the formal competition outcomes match the knowledge I have gained from this course.” Each item in the second part is measured by a nine-point Likert scale ranging from 1 (extremely disagree) to 9 (extremely agree). The questionnaire items stated above are presented in the later section of Data Analysis, along with the following statistical methods.
Statistical methods

Three different statistical methods are used in this study. First, descriptive analysis is used to describe the profile of the sample students and to inspect visually the measurement items and their relationships in regards to the three research propositions. Second, repeated measures ANOVA is used to test formally the differences in students’ perceived complexity level as well as the skill and knowledge gains among the three BSGs, after the Mauchly’s test of sphericity. When significant differences are found in the ANOVA, Bonferroni-corrected pairwise comparisons are applied to further identify which pair of BSGs contributes to the significant differences. Repeated measures ANOVA is used because the same students provide ratings on the measures on all three games. Third, paired sample t-test is employed to test the perceived gains in knowledge and skills after each of the five activity stages during the learning cycles. Since this research intends to provide insights on the incremental gains of the same students between two consecutive stages among the five activities in the leaning cycles, the paired sample t-test is adopted. To be specific, this research is only interested in finding significant mean differences of students’ skills and knowledge between two consecutive stages (i.e., stage 1 versus 2, 2 versus 3, 3 versus 4, and 4 versus 5), not all stages pairs (e.g., stage 1 versus 5). SPSS statistical software is used to analyze the collected data with the statistical methods mentioned above.

Results and discussion

Sample profile and descriptive statistics

The general course was offered to all undergraduate students in a national university in Taiwan. More than 150 students pre-registered for the general course Business Simulation Games and Competition; however, only 51 were officially approved by the online registration system. Among the 51 students, 43 completed the questionnaire administered at the end of the class.

A brief summary of the sample profile is as follows. Among the student respondents, 67.4% were male. Students majoring in Engineering, Management, Science, Law, and Humanities displayed a distribution of 37.2%, 30.2%, 18.6%, 9.3%, and 4.7%, respectively. Although students majoring in management only comprise 30.2% of the respondents, more than half (51.16%) of the participants have taken management course before and only 25.6% of the students have experiences using BSGs in previous classes. In terms of average scores in the previous semester, majority of the respondents obtained an average of 80s (55.8%) and 70s (34.9%), whereas only a few received an average of 50s, 60s, and 90s. This student profile fulfills our intended research setting; that is, an appropriate mixture of diverse student backgrounds.

Table 1 summarizes the measurement items for the perceived complexity level of three BSGs, the perceived gains in skills and three types of knowledge, and the perceived match between BSG competition outcomes and their gains in skills and knowledge. Table 1 also provides the mean and standard deviation for each BSG.

As shown in Table 1, the mean falls between 2.34 and 7.29. The score distribution is relatively even for the gains in knowledge and skills at different leaning activity stages, whereas the distribution is relatively different for the complexity level on the basis of our visual inspection.

To assess the different complexity levels of the three BSGs, the means of item C1 “the game is complicated” as perceived by the 43 students were used. Given that the means for Beer Game, Retail Expert, and BOSS are 3.59, 5.29, and 7.14, respectively, the assessment supports the perceived difference in the complexity level. A reverse item C2 “the game is easy to use” was used to confirm the statement of C1. The means for Beer Game, Retail Expert, and BOSS are 7.29, 6.28, and 4.81, respectively, indicating a reverse trend to that of C1. Therefore, the levels of game complexity fulfill the intended research setting; that is, the three BSGs represent different complexity levels.

The means of the remaining items for the perceived gains in skills and three types of knowledge show similar ranges and trends across three BSGs. For instance, the means of K1 in skills range from 2.52 to 2.82, whereas the means of K5 range from 6.53 to 6.70. Similarly, the means of D1, P1, and S1 in the three different types of knowledge range
The means of D5, P5, and S5 range from 5.72 to 6.44. This result indicates that the gains in skills and three types of knowledge at different stages remain consistent even when learning BSGs across three different levels of complexity; thus, proposition 1 seems initially confirmed.

<table>
<thead>
<tr>
<th>Item</th>
<th>Item</th>
<th>Beer Game</th>
<th>Retail Expert</th>
<th>BOSS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean(^a)</td>
<td>Mean(^a)</td>
<td>Mean(^a)</td>
</tr>
<tr>
<td>Complexity level</td>
<td></td>
<td>SD(^b)</td>
<td>SD(^b)</td>
<td>SD(^b)</td>
</tr>
<tr>
<td>C1. The game is complicated</td>
<td>C2. The game is easy to use</td>
<td>3.59</td>
<td>1.51</td>
<td>5.29</td>
</tr>
<tr>
<td>Skills</td>
<td></td>
<td>7.29</td>
<td>1.16</td>
<td>6.28</td>
</tr>
<tr>
<td>K1. Before operating the BSG</td>
<td>K2. After operating the BSG</td>
<td>2.82</td>
<td>2.16</td>
<td>2.84</td>
</tr>
<tr>
<td>K3. After practicing the BSG</td>
<td>K4. After writing the group thought-sharing report</td>
<td>6.25</td>
<td>1.37</td>
<td>5.57</td>
</tr>
<tr>
<td>K5. After completing the final game</td>
<td></td>
<td>6.93</td>
<td>1.47</td>
<td>6.53</td>
</tr>
<tr>
<td>Declarative knowledge</td>
<td></td>
<td>3.59</td>
<td>1.87</td>
<td>2.64</td>
</tr>
<tr>
<td>D1. Before operating the BSG</td>
<td>D2. After operating the BSG</td>
<td>4.18</td>
<td>1.63</td>
<td>4.30</td>
</tr>
<tr>
<td>D3. After practicing the BSG</td>
<td>D4. After writing the group thought-sharing report</td>
<td>5.11</td>
<td>1.47</td>
<td>4.77</td>
</tr>
<tr>
<td>D5. After completing the final game</td>
<td></td>
<td>5.93</td>
<td>1.53</td>
<td>5.72</td>
</tr>
<tr>
<td>Procedural knowledge</td>
<td></td>
<td>2.56</td>
<td>1.83</td>
<td>2.63</td>
</tr>
<tr>
<td>P1. Before operating the BSG</td>
<td>P2. After operating the BSG</td>
<td>4.23</td>
<td>1.64</td>
<td>4.12</td>
</tr>
<tr>
<td>P3. After practicing the BSG</td>
<td>P4. After writing the group thought-sharing report</td>
<td>5.12</td>
<td>1.42</td>
<td>4.95</td>
</tr>
<tr>
<td>P5. After completing the final game</td>
<td></td>
<td>6.28</td>
<td>1.64</td>
<td>5.88</td>
</tr>
<tr>
<td>Strategic knowledge</td>
<td></td>
<td>2.36</td>
<td>1.81</td>
<td>2.43</td>
</tr>
<tr>
<td>S1. Before operating the BSG</td>
<td>S2. After operating the BSG</td>
<td>3.66</td>
<td>1.92</td>
<td>3.86</td>
</tr>
<tr>
<td>S3. After practicing the BSG</td>
<td>S4. After writing the group thought-sharing report</td>
<td>5.23</td>
<td>1.58</td>
<td>5.00</td>
</tr>
<tr>
<td>S5. After completing the final game</td>
<td></td>
<td>6.41</td>
<td>1.37</td>
<td>6.12</td>
</tr>
<tr>
<td>Matching the competition</td>
<td></td>
<td>6.05</td>
<td>1.43</td>
<td>5.90</td>
</tr>
<tr>
<td>M1. The competition outcome matches the overall knowledge gains</td>
<td>5.88</td>
<td>1.71</td>
<td>5.95</td>
<td>1.45</td>
</tr>
</tbody>
</table>

\(^a\)A nine-point Likert scale, where 1 represents totally disagree and 9 represents totally agree. \(^b\)Standard deviation.

Finally, the perceived match between the competition outcomes and gains in knowledge (M1) and skills (M2) are from slightly positive to mildly positive, sufficing proposition 3 and supporting the IPO model of Garris et al. (2002) and the PBG model of Kiili (2007).
Inferential statistical tests

In addition to inspecting the descriptive statistics, the formal statistical procedures of repeated measures ANOVA and paired sample t-test are adopted to examine the research questions. Before applying the repeated measures ANOVA, the sphericity assumptions of the measures are tested first. Table 2 shows the results of the Mauchly's test of sphericity.

<table>
<thead>
<tr>
<th>Item</th>
<th>Complexity level</th>
<th>Skills</th>
<th>Declarative knowledge</th>
<th>Procedural knowledge</th>
<th>Strategic knowledge</th>
<th>Matching the competition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mauchly's W</td>
<td>Mauchly's W</td>
<td>Approx.</td>
<td>Degree of freedom</td>
<td>Approx. Chi-Square</td>
<td>Degree of freedom</td>
<td>Approx. Chi-Square</td>
</tr>
<tr>
<td>C1. The game is complicated</td>
<td>.95</td>
<td>2.15</td>
<td>2</td>
<td>.342</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C2. The game is easy to use</td>
<td>.93</td>
<td>2.98</td>
<td>2</td>
<td>.226</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1. Before operating the BSG</td>
<td>.99</td>
<td>2.60</td>
<td>2</td>
<td>.269</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K2. After operating the BSG</td>
<td>.94</td>
<td>2.34</td>
<td>2</td>
<td>.307</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K3. After practicing the BSG</td>
<td>.84</td>
<td>7.02</td>
<td>2</td>
<td>.030**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M1. The competition outcome matches the</td>
<td>.86</td>
<td>6.25</td>
<td>2</td>
<td>.044***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>overall knowledge gains</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M2. The competition outcome matches the</td>
<td>.92</td>
<td>3.50</td>
<td>2</td>
<td>.174</td>
<td></td>
<td></td>
</tr>
<tr>
<td>overall skill gains</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Mauchly's test of sphericity

The significant p-values of the measures of K1, D1, D3, D4, P1, P3, P4, S1, and M2 indicate a violation of the sphericity assumption, and thus corrections are used when testing significance in the repeated measures ANOVA. We adopt the Greenhouse-Geisser correction for K1 because the Mauchly’s test of sphericity shows a Greenhouse-Geisser epsilon of < .75, and the Huynh-Feldt correction for D1, D3, D4, P1, P3, P4, S1, and M2 because of a
Greenhouse-Geisser epsilon of > .75 (Girden, 1992). Table 3 shows the results from the repeated measures ANOVA comparing the three BSGs.

Table 3. Repeated measures ANOVA results comparing the three BSGs

<table>
<thead>
<tr>
<th>Item</th>
<th>Type III Sum of squares</th>
<th>Degrees of freedom</th>
<th>Mean square</th>
<th>F value</th>
<th>p valuea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complexity level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1. The game is complicated</td>
<td>257.29/134.71</td>
<td>2/82</td>
<td>128.64/1.64</td>
<td>78.30</td>
<td>.000***</td>
</tr>
<tr>
<td>C2. The game is easy to use</td>
<td>128.05/138.62</td>
<td>2/82</td>
<td>64.02/1.69</td>
<td>37.87</td>
<td>.000***</td>
</tr>
<tr>
<td>Skills</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K1. Before operating the BSG</td>
<td>2.06/113.94</td>
<td>1.37/57.39</td>
<td>1.51/1.99</td>
<td>.76</td>
<td>.426</td>
</tr>
<tr>
<td>K2. After operating the BSG</td>
<td>3.27/100.06</td>
<td>2/82</td>
<td>1.64/1.19</td>
<td>1.37</td>
<td>.259</td>
</tr>
<tr>
<td>K3. After practicing the BSG</td>
<td>9.78/122.22</td>
<td>2/82</td>
<td>4.89/1.46</td>
<td>3.36</td>
<td>.039*</td>
</tr>
<tr>
<td>K4. After writing the group thought-sharing report</td>
<td>1.50/55.16</td>
<td>2/84</td>
<td>.75/.66</td>
<td>1.15</td>
<td>.323</td>
</tr>
<tr>
<td>K5. After completing the final game</td>
<td>2.99/107.67</td>
<td>2/84</td>
<td>1.50/1.28</td>
<td>1.17</td>
<td>.316</td>
</tr>
<tr>
<td>Declarative knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D1. Before operating the BSG</td>
<td>1.60/38.40</td>
<td>1.69/71.13</td>
<td>.94/54</td>
<td>1.75</td>
<td>.186</td>
</tr>
<tr>
<td>D2. After operating the BSG</td>
<td>5.04/81.63</td>
<td>2/84</td>
<td>2.52/97</td>
<td>2.59</td>
<td>.081</td>
</tr>
<tr>
<td>D3. After practicing the BSG</td>
<td>2.28/93.05</td>
<td>1.73/72.58</td>
<td>1.32/1.28</td>
<td>1.03</td>
<td>.353</td>
</tr>
<tr>
<td>D4. After writing the group thought-sharing report</td>
<td>.57/80.09</td>
<td>1.80/75.41</td>
<td>.32/1.06</td>
<td>.30</td>
<td>.717</td>
</tr>
<tr>
<td>D5. After completing the final game</td>
<td>1.72/88.95</td>
<td>2/84</td>
<td>.86/1.06</td>
<td>.81</td>
<td>.447</td>
</tr>
<tr>
<td>Procedural knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1. Before operating the BSG</td>
<td>1.21/24.13</td>
<td>1.73/70.81</td>
<td>.70/34</td>
<td>2.05</td>
<td>.143</td>
</tr>
<tr>
<td>P2. After operating the BSG</td>
<td>.25/65.75</td>
<td>2/82</td>
<td>.13/80</td>
<td>.16</td>
<td>.854</td>
</tr>
<tr>
<td>P3. After practicing the BSG</td>
<td>1.48/86.52</td>
<td>1.73/70.89</td>
<td>.85/1.22</td>
<td>.70</td>
<td>.480</td>
</tr>
<tr>
<td>P4. After writing the group thought-sharing report</td>
<td>.68/46.65</td>
<td>1.77/72.50</td>
<td>.39/.64</td>
<td>.60</td>
<td>.532</td>
</tr>
<tr>
<td>P5. After completing the final game</td>
<td>3.11/71.56</td>
<td>2/84</td>
<td>1.56/87</td>
<td>1.78</td>
<td>.175</td>
</tr>
<tr>
<td>Strategic knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1. Before operating the BSG</td>
<td>.11/31.23</td>
<td>1.77/74.15</td>
<td>.06/42</td>
<td>.15</td>
<td>.839</td>
</tr>
<tr>
<td>S2. After operating the BSG</td>
<td>.61/84.73</td>
<td>2/84</td>
<td>.30/1.01</td>
<td>.30</td>
<td>.742</td>
</tr>
<tr>
<td>S3. After practicing the BSG</td>
<td>1.69/98.98</td>
<td>2/84</td>
<td>.85/1.18</td>
<td>.72</td>
<td>.491</td>
</tr>
<tr>
<td>S4. After writing the group thought-sharing report</td>
<td>5.21/95.46</td>
<td>2/84</td>
<td>2.61/1.14</td>
<td>2.29</td>
<td>.107</td>
</tr>
<tr>
<td>S5. After completing the final game</td>
<td>1.81/122.85</td>
<td>2/84</td>
<td>.91/1.46</td>
<td>.62</td>
<td>.540</td>
</tr>
<tr>
<td>Matching the competition.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M1. The competition outcome matches the overall knowledge gains</td>
<td>.98/104.36</td>
<td>2/84</td>
<td>.49/1.24</td>
<td>.39</td>
<td>.676</td>
</tr>
<tr>
<td>M2. The competition outcome matches the overall skill gains</td>
<td>2.53/138.14</td>
<td>1.82/76.54</td>
<td>1.39/1.81</td>
<td>.77</td>
<td>.456</td>
</tr>
</tbody>
</table>

a*** p < 0.001. ** p < 0.01. * p < 0.05.

The basic null hypothesis $H_0$ assumes equal perceived means of the three BSGs. The $p$ values for complexity (C1) and simplicity (C2) are both significant at the 0.001 level; thus, the null hypotheses of equal means are rejected. This finding confirms our design to have three BSGs with different levels of complexity. By contrast, except K3 which is
significant at the 0.05 level, all of the items for the gains in skills and three knowledge categories across the three BSGs have the same ANOVA results of accepting the null hypotheses of equal means. The Bonferroni-corrected pairwise comparisons for K3 reveal a significant skill level difference between the Beer Game (M = 6.21, SD = 1.36) and the Retail Expert (M = 5.53, SD = 1.56).

According to the summary shown in Table 3, the ANOVA analyses results partially support proposition 1. Therefore, although the three BSGs have different complexity levels, the student respondents perceived the BSGs more or less similar in improving the levels of skills, declarative knowledge, procedural knowledge, and strategic knowledge. Furthermore, the test results of M1 and M2 are also insignificant; that is, the students do not perceive the level of match between the competition outcomes and their overall knowledge and skill gains significantly different across the three BSGs in this course. Hence, proposition 3 is supported.

This study focuses on learning how skills and knowledge are incrementally gained at different stages of the BSG learning cycles. The BSG learning cycle activities are designed into five stages in this study. Therefore, the perceived incremental gain is defined as the significant mean difference of perceived gains between two consecutive stages, including the gains after operational instruction, after informal practicing, after the group has written the thought-sharing report, and after the formal competition. Therefore, the paired sample t-test is applied to test the significance of the differences, as summarized in Table 4.

On the basis of Table 4, several observations can be derived. First, for all the four categories of skills and knowledge, all but one (i.e., except Retail Expert) incremental gains across stages are significant after the introduction and informal practice of BSGs. Therefore, providing a good BSG introduction to learners is important to achieve greater gains from the beginning of the instruction.

Second, for the four stages, the gains in either skills or the types of knowledge show different patterns; however, these gains demonstrate a plausible potential for significant incremental increase at different stages for different BSGs. Therefore, the appropriateness of the repeating cycles in the IPO game model of Garris et al. (2002) is confirmed. An in-depth discussion of each of the skills and knowledge is provided in our observations 3 to 6.

Third, the gains after the K4-K3 skills of written thought-sharing report depend on the individual BSG. In this case, Retail Expert would continuously increase throughout the entire learning cycles. By contrast, the Beer Game stopped after K3-K2, that is, after informal practice; BOSS stopped after K4-K3, that is, after the thought-sharing written report. This result may be caused by the differences in the decision items of the games; specifically, Retail Expert and BOSS have more decision items to manage, whereas Beer Game only needs to enter the order for the next period. Therefore, decision complexity may be a critical factor for gaining the operational skills of BSGs.

Fourth, the Beer Game and BOSS are both strong in gaining declarative knowledge throughout the stages. By contrast, gains after practicing (D3-D2) and after completing the final game (D5-D4) are not significant in Retail Expert. Retail Expert is a personal game that requires intensive practice to become competitive; therefore, a player may lose his/her competitive edge if his/her rivals gain more declarative knowledge after intensive practices. Therefore, the game Retail Expert results in students not feeling confident in their accumulated declarative knowledge when losing games during practice and formal competition. However, the group thought-sharing report helps the students gain confidence in their accumulated declarative knowledge, given that this report is a brainstorming-like discussion within the group that summarizes the declarative knowledge of students. This report also provides students the opportunity to see the reports of other groups. Therefore, declarative knowledge should increase throughout the repeating cycles despite the exception of the Retail Expert instance, as explained above.

Fifth, procedural knowledge, except for that of the Beer Game, seems to stop after P5-P4, which is the stage after the written thought-sharing report. This result may imply that procedural knowledge will not increase after a certain stage because the operators of the process may be exhausted after a certain number of cycles in the game, even in the presence of new declarative or strategic knowledge.

Sixth, the incremental gains of strategic knowledge seem to be the strongest because of the repeating cycles among these four categories, except for the Beer Game. As indicated by Kiili’s model, strategies can be changed during competition cycles and can be versatile because of the dynamic competition environment. Therefore, continuously increasing strategic BSG knowledge during the leaning cycles is reasonable.
Utilization of the learning activities to gain the three knowledge categories may be further affected by different complexity levels in game practices for accumulating certain experiences and skills. Therefore, teachers should at least arrange two to three activities in a row to provide students with significant gains in related knowledge and skills. Second, procedural knowledge gain may be maximized after group discussion depending on the experiences of the participants because procedural knowledge is similar to playing skills that connect the dots (i.e., declarative knowledge). Therefore, to obtain the maximum effect of BSGs, the teachers should ensure that an adequate number of game practices for accumulating certain experiences are provided before the students proceed to group-discussion activities. Third, both strategic and declarative knowledge may have more chances to grow continuously with increasing cycles of learning activities. In a dynamic game-playing environment, declarative knowledge may not be discovered when the corresponding scenarios are not encountered and when strategies can be flexibly changed to influence the competition outcomes. Therefore, the perceived knowledge of these two categories may be further addressed with more activities during the learning cycles, as suggested by Garris et al. (2002) and Kiili (2007). Finally, certain exceptions or uncertainty may exist with different complexity levels or characteristics of the BSGs; for instance, the evidence for the gains in skill and strategic knowledge when playing Retail Expert is strong, but the

### Table 4. Paired sample t-test analysis of the mean differences

<table>
<thead>
<tr>
<th>Item</th>
<th>Beer game</th>
<th>Retail expert</th>
<th>BOSS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean Difference&lt;sup&gt;a&lt;/sup&gt;</td>
<td>p value&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Mean Difference&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Skills</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K2-K1. gains after operating BSGs</td>
<td>2.20</td>
<td>.000***</td>
<td>2.11</td>
</tr>
<tr>
<td>K3-K2. gains after practicing BSGs</td>
<td>1.23</td>
<td>.000***</td>
<td>.61</td>
</tr>
<tr>
<td>K4-K3. gains after writing thought-sharing report</td>
<td>.30</td>
<td>.208</td>
<td>.70</td>
</tr>
<tr>
<td>K5-K4. gains after completing the final game</td>
<td>.39</td>
<td>.101</td>
<td>.28</td>
</tr>
<tr>
<td><strong>Declarative knowledge</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D2-D1. gains after operating BSGs</td>
<td>1.61</td>
<td>.000***</td>
<td>1.66</td>
</tr>
<tr>
<td>D3-D2. gains after practicing BSGs</td>
<td>.93</td>
<td>.000***</td>
<td>.48</td>
</tr>
<tr>
<td>D4-D3. gains after writing thought-sharing report</td>
<td>.52</td>
<td>.004**</td>
<td>.80</td>
</tr>
<tr>
<td>D5-D4. gains after completing the final game</td>
<td>.30</td>
<td>.026*</td>
<td>.12</td>
</tr>
<tr>
<td><strong>Procedural knowledge</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P2-P1. gains after operating BSGs</td>
<td>1.67</td>
<td>.000***</td>
<td>1.49</td>
</tr>
<tr>
<td>P3-P2. gains after practicing BSGs</td>
<td>.88</td>
<td>.000***</td>
<td>.84</td>
</tr>
<tr>
<td>P4-P3. gains after writing thought-sharing report</td>
<td>.58</td>
<td>.009**</td>
<td>.56</td>
</tr>
<tr>
<td>P5-P4. gains after completing the final game</td>
<td>.58</td>
<td>.009**</td>
<td>.33</td>
</tr>
<tr>
<td><strong>Strategic knowledge</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2-S1. gains after operating BSGs</td>
<td>1.30</td>
<td>.000***</td>
<td>1.43</td>
</tr>
<tr>
<td>S3-S2. gains after practicing BSGs</td>
<td>1.57</td>
<td>.000***</td>
<td>1.14</td>
</tr>
<tr>
<td>S4-S3. gains after writing thought-sharing report</td>
<td>.89</td>
<td>.000***</td>
<td>.66</td>
</tr>
<tr>
<td>S5-S4. gains after completing the final game</td>
<td>.30</td>
<td>.057</td>
<td>.47</td>
</tr>
</tbody>
</table>

<sup>a</sup>Mean difference of two learning activities. <sup>b</sup>p < 0.001. <sup>*</sup>p < 0.01. <sup>**</sup>p < 0.05.

### Conclusions and implications

With the above data analyses, proposition 3 is fully supported, whereas proposition 1 and 2 are partially sustained. This research contributes in the validation of the concept of BSG learning cycles in the theoretical models of Garris et al. (2002) and Kiili (2007). The learning-cycle effect and the utilization of the learning activities to gain the three types of knowledge and skills have implications as well as limitations.

Four implications can be derived in this study. First, the first few fundamental activities are critical in quickly gaining substantial skills and knowledge when playing BSG. Thus, teachers planning to adopt BSGs should at least arrange two to three activities in a row to provide students with significant gains in related knowledge and skills. Second, procedural knowledge gain may be maximized after group discussion depending on the experiences of the participants because procedural knowledge is similar to playing skills that connect the dots (i.e., declarative knowledge). Therefore, to obtain the maximum effect of BSGs, the teachers should ensure that an adequate number of game practices for accumulating certain experiences are provided before the students proceed to group-discussion activities. Third, both strategic and declarative knowledge may have more chances to grow continuously with increasing cycles of learning activities. In a dynamic game-playing environment, declarative knowledge may not be discovered when the corresponding scenarios are not encountered and when strategies can be flexibly changed to influence the competition outcomes. Therefore, the perceived knowledge of these two categories may be further addressed with more activities during the learning cycles, as suggested by Garris et al. (2002) and Kiili (2007). Finally, certain exceptions or uncertainty may exist with different complexity levels or characteristics of the BSGs; for instance, the evidence for the gains in skill and strategic knowledge when playing Retail Expert is strong, but the
gains in declarative knowledge are rather weak. Therefore, choosing the appropriate complexity level of BSG significantly contributes to the learning effect of the activity.

The research outcomes provide several insights that would help teachers relate class activities with the learning-cycle effects in terms of student gains in the three types of knowledge and skills. Nevertheless, two research limitations exist. First, the sample size can be expanded to cover more classes in different universities in Taiwan to generalize the research findings from a small sample size to a larger applicable scope, such as the entire Taiwan. Second, a greater variety of class activities can be added in future research for more drill-down analyses, thus providing teachers more insightful information regarding which activities may be more appropriate or effective at certain scenarios.

References


A Learning Cycle Approach to Developing Educational Computer Game for Improving Students’ Learning and Awareness in Electric Energy Consumption and Conservation

Ugyen Dorji¹, Patcharin Panjaburee²* and Niwat Srisawasdi²

¹Institute for Innovative Learning, Mahidol University, Thailand // ²Faculty of Education, Khon Kaen University, Thailand // ugyend505@gmail.com // panjaburee_p@hotmail.com // niwswi@kku.ac.th

*Corresponding author

(Submitted January 29, 2014; Revised June 22, 2014; Accepted July 8, 2014)

ABSTRACT

Educating on the knowledge of electric energy consumption and conservation has been becoming a focus and rather urgent issue in energy education. Although several researchers have attempted to apply teaching and learning strategies to promote students’ conceptual understanding on the topic, they still have sub-optimal conceptual understanding of what the major factors of energy consumption are. This led to less awareness of electric energy conservation. With the benefit of educational computer game in motivating students’ learning, this study developed a computer game based on a learning cycle approach for assisting learning on the topic. When using the computer game, the students were encouraged to use electrical appliances for certain durations. To examine the effectiveness of the computer game, the study was conducted with 129 tenth grade students in eastern Bhutan. The experimental results showed that the computer game significantly improved the students’ learning achievement on energy consumption as well as their awareness on electric energy conservation.

Keywords

Electricity, Energy consumption, Inquiry, Game-based learning, Self-awareness

Introduction

With the significance of saving energy, researchers are deeply aware of the importance of teaching and learning energy concept in classrooms. A great deal of effort has put into research related to the development of energy educational program to facilitate students to identify the effective way to save energy (DeWaters & Powers, 2011). Nevertheless, some researchers showed that such programmes often were not put into actions (Vastamäki, Sinkkonen, & Leinonen, 2005). Moreover, previous research indicated that most people have only a vague idea of how much electricity is consumed by their household appliances or do not have any idea of how to save electricity in an appropriate manner, and what sort of difference they could make by changing day-to-day behavior or investing in efficient measures. For this reason, energy consumption has not decreased significantly in reality (Boyd, 2002). Therefore, teaching them the basic concepts, such as calculating energy consumption by each of their household electrical appliances and what would be the approximate monthly electric bill, how to save unnecessary waste of energy by providing feedback at the right moment, are necessary for developing cautiousness in energy usage. Moreover, the adaptive energy saving tips could help students learn to adopt appropriate levels of electricity consumption (Haakana, Sillanpää, & Talsi, 1997). In this view, the importance of energy consumption and conservation concept in physics or in energy education several researchers have attempted to promote students’ conceptual understanding to equip with appropriate ways to educate this concept for a secondary school level. For example, Gustafson and Branch (2002) designed instructional models which fit varied learning situation and introduced the concept of electrical energy consumption by explaining in an effective way. Bates (2003) applied the methods of instructional technology to serve the concept. However, students still lack the knowledge of a daily life electrical appliance and they have few conceptual understanding about the major factors of energy consumption. Therefore, there is little knowledge and awareness on energy conservation.

With the rapid growth of technology-enhanced learning approach, educational computer game has been recognized as to motivate students’ learning in several courses. In recent years, there have been many developments of computer games that were used for supporting learning and teaching in different levels of education. For example, Cai, Lu, Zheng, and Li (2006) developed an immersive protein gaming for biology course. Virvou, Katsionis, and Manos (2005) designed a virtual reality games to teach geography students. Connolly, Stansfield, and Hainey (2007)
developed computer game for a software engineering course. Yien, Hung, Hwang, and Lin (2011) proposed a game-based learning approach in improving students’ learning achievements in a nutrition course. Yang, Chien, and Liu (2012) developed a digital game-based learning system for energy education for the children in the school. These studies have pointed out that the educational computer game could engage the students in meaningful motivational learning. But on other hand, without incorporating educational theory or applying teaching and learning strategies, the educational computer games might not be significant for a meaningful learning (Ke, 2008). Consequently, this study attempts to apply the learning cycle approach as an educational-theoretical framework for developing an educational computer game, in which students are able to explore and explain the situation occurring in the developed game for constructing electric energy consumption and conservation knowledge. Meanwhile, the developed game characteristics are utilized to promote their awareness on energy conservation.

To our knowledge, there is no research concerning about the integration of multimedia into the 5E learning cycle-based activities for improving students’ learning and awareness in electrical energy consumption and conservation. With the benefit of the 5E learning cycle model and the educational computer game, this study incorporated such learning approach into the educational computer game serving as supplementary material after a traditional lecture of abstract concepts of household electrical energy consumption and conservation. An experiment was conducted in a secondary school Physics course “electric energy consumption and conservation” to investigate the following research questions:

- Do the students who learn with the educational game based on a learning cycle approach have better learning achievement than those who learn with the conventional teaching approach?
- Do the students who learn with the educational game based on a learning cycle approach have better awareness toward energy conservation than those who learn with the conventional teaching approach?
- How is the students’ satisfaction toward the educational game based on a learning cycle approach?

Literature review

Educational computer game

Computer games play an important role in any fields of education. The use of computer games in education has shown positive learning outcomes by using it as a tool for instructional purposes. Several studies showed that the use of educational computer games increase motivation and better learning achievement. The positive impact was well received as new approach in teaching by many students when compared with traditional teaching. For example, Sofoluwe (2007) stated that educational games are good tools to promote students participation in learning activities and then to motivate them in their learning as it encourages learning within enjoyable environment. Through game play, teacher can encourage student’s abstract thinking into a concrete concept and further foster their higher order thinking abilities. Computer games are able to boost learning motivation of players through its features like adventures, challenges and fantasies. Moreover, playing educational computer game for long-term had a positive effect on students’ learning abilities and construct knowledge (Burguillo, 2010; Liu, Peng, Wu, & Lin, 2009). Computer games are claimed to have effects on cognitive development through visual skills including “spatial representation,” “iconic skill,” and “visual attention.” While playing games the player(s) become more skilled in games and their visual attention becomes proportionally better. Moreover, Dondlinger (2007) articulated that motivation by using game can promote the learning achievement even in the complex concept. For example, a study of Ke (2008) indicated that the educational computer games had more impact in promoting motivation in learning than conventional chalk and talk method.

In general, games are employed to motivate students to take responsibility for their own learning that leads to intrinsic motivation. Prensky (2011) indicated that gaming activities had potential to engage the students into a state of flow and consequently cause better learning through focus and pleasant rewards while increasing their motivation and attainment. Consequently, in recent years numerous educational computer games have been designed in several difficult subject domains to supplement teaching and learning for improving students’ academic achievement, including mathematics (Lee & Chen, 2009), language (Chen, Liao, Chien, & Chan, 2011), software programming (Connolly et al., 2007), management education (Chen et al., 2013; Kiili, 2007), medical education (Moreno-Ger, Burgos, Martínez-Ortiz, Sierra, & Fernández-Manjón, 2008), and sports education (Mueller, Gibbs, & Vetere, 2010).
Educational computer games have been used to support active, productive, creative and collaborative learning methods (Hoppe, Joiner, Milrad, & Sharples, 2003). However, the integration of the computer game approach into the classroom, teaching does not necessarily ensure successful learning. Therefore, to ensure effective and successful learning, development of any innovative approach need to incorporate appropriate learning strategies that fit both in the learners and to the teachers (Charsky & Ressler, 2011; Chuang & Chen, 2009; Wang & Chen, 2010).

A learning cycle approach

The 5E learning cycle model-oriented learning cycle approach is a realistic, constructivist method of learning which employs students through well-designed learning process (Bybee, 2004). It is based on the constructivism theory of John Dewey and Jean Piaget. It was first designed by Robert Karplus in the early 1960s, and later in 2004, Bybee conducted full study and developed a method, named 5E Learning Cycle model. According to Bybee (2004), the 5E learning cycle model consists of five phase inquiry approaches: (1) Engagement: In this phase the teacher or a curriculum task accesses the students’ prior knowledge and engages them in a new concept through the use of short activities that promote curiosity and stimulate prior knowledge; (2) Exploration: During this phase students are provided opportunity to experience activities within current concepts, skills and processes. In this phase, conceptual change is facilitated with guidance from teacher/instructor; (3) Explanation: This phase focuses students’ attention on a particular aspect of their engagement and exploration experiences and provides opportunities to share and explain their understanding of the concepts, skills and processes that they have learned. Here, it also provides opportunities for teachers to clarify students’ misconceptions and help them in introducing new concepts, process and skills; (4) Elaboration: As they received theoretical concepts, process and skills, now they need to experience further related to new situation. This phase enable students to extend the concepts they have learned, link to other related concepts and apply their conceptual understandings into the real world situations in a newer ways. Moreover, teacher further challenges and extends students’ conceptual understanding and skills of the concept by conducting additional activities that focuses on adding breadth and depth to current understanding; and (5) Evaluation: The evaluation phase provides students’ opportunities to self-assessment as well as formal assessment for both students and instructors to determine how much learning and conceptual understanding has taken place and whether they have met the learning outcomes.

Previous studies advocated that the instruction based on the 5E learning cycle model had positive effect on students’ achievement. Moreover, the 5E learning cycle instruction was able to relate their conceptual learning with ability to apply in their daily life situations, which was the gap in that of traditional learning method (Kaveevivitchai et al., 2009; Kaynar Tekkaya, & Cakiroğlu, 2009; Liu et al., 2009). For example, Kaynar et al. (2009) showed that the 5E learning cycle model was more effective in enhancing students’ achievement in cell concepts topic over the traditional learning. Beside the 5E Learning Cycle model has been recognized as one of the best teaching and learning approach in learning abstract concepts in several subjects in school education (Cepni & Sahin, 2012; Tuna & Kacar, 2013). For example, Cepni and Sahin (2012) concluded that students who learned with the instruction based on the 5E learning cycle model had better learning outcomes than those who learned with other teaching methods in learning buoyancy force. A recent study by Tuna and Kacar (2013) indicated teaching trigonometry concepts with 5E learning cycle model enhanced students’ academic achievement and longevity of their knowledge. Therefore, the aforementioned researches suggest the 5E learning cycle model is one of the widely-adopted pedagogies that involve brain-hands-mind in learning for enhancing students’ conceptual understanding. This model is a guided inquiry-based scientific pedagogy (Bybee, 2004) where students are engaged in new concept by relating their previous knowledge by exploring and explaining through their experiences, then elaborating on what they have learned, and eventually evaluating their understanding on that new concept under the guidance of teachers (Uzunöz, 2011). When organized and correctly operated, this 5E learning cycle model enhances not only the students’ achievement but also the permanence of knowledge in various fields of education.
Development of an educational computer game based on the 5E learning cycle model for energy consumption and conservation

With the benefit of the 5E learning cycle model and the educational computer game, this study incorporated such learning model into the educational computer game serving as supplementary material after a traditional lecture of abstract concepts of household electrical energy consumption and conservation. The developed educational computer game has scoped to make students understand the basic concept of daily household energy usage by identifying the basic factors that determine energy consumption, calculation of energy bill and develop awareness on energy conservation to save money in meaningful manners.

The procedure of conducting the learning activities in the developed educational computer game is shown in Figure 1. It consists of six stages as follows:

Stage 1: The objectives of the learning activities are introduced to the students, such as the students should be able to identify the factors that effect on electrical energy consumption in electrical appliances; calculate the energy consumption and cost for operating daily used household electrical appliances; and inculcate better sense of awareness on the energy conservation. Before participating in the learning activities, they are asked to check their prior knowledge on electric energy consumption. Moreover, the rules and basic functions of the game are also demonstrated.

Stage 2: The students receive the simulated situation in which they are encouraged to investigate the cause of power consumption. The teacher will guide them with questions, for example what kinds of electrical appliances operated at your home and how much energy is consumed by those appliances. In this stage, teacher will link students’ prior knowledge to the learning activities.

Stage 3: The students receive the first computer game in which they are asked to investigate the energy consumption factors in electrical appliances. This game is divided into two cases. In the first case, the students are asked to select electrical appliance with different wattage and then select the time to compare the energy consumption. Suddenly, the graph of energy consumption will be displayed based upon selected appliances with selected time. The system will guide them to observe the displayed graph to identifying the first factor “wattage” for energy consumption in appliances, as shown in Figure 2(a). Similarly, in the second case, the students are asked to select time for each electrical appliance which has the same wattage to compare the energy consumption. Suddenly, the graph of energy consumption will be displayed based upon selected appliances with different time. The system will guide them to observe the displayed graph to identifying the second factor “time/duration” for energy consumption in appliances, as shown in Figure 2(b).

Stage 4: The students are asked to summarize their finding from Stage 3. The system will guide them to explain the relationship among wattage, voltage, current and time. If they explain incorrectly, the system will show hints and guide them with the displayed graph in Stage 3 once again. Moreover, the teacher will provide meaning of electrical appliance rating, formula for calculating energy consumption of each appliance and also the monthly electric bill in more details.

Stage 5: The students receive the second computer game for applying the constructed knowledge from Stage 4 to new situation. The game is provided with the scenario of home comprising of electrical appliances. Here students are asked to use the electrical appliance(s) with the time limitation for saving money on power bill as much as they can, as shown in Figure 3. If the students fail to save money meaning that they pay much than their income, the system will provide them with hints about the detailed factors of energy consumption.

Stage 6: The students receive the third computer game, that is, to evaluate their own conceptual understanding about energy consumption and awareness of energy conservation, as shown in Figure 4. Once they have finished shopping electrical appliances with the time limitation, the system will show the game result regarding to the different ways for making the appliances more energy efficient.
Stage 1: Learning activity Orientation
- Introduce the learning objectives
- Check the prior knowledge on electric energy consumption
- Demonstrate the rules and basic function of the game

Stage 2: Engage students’ learning in electrical energy consumption and conservation
- Provide scenario-based happenings with the electrical bill;
- Encourage students to list down the electrical appliances operated at their home;
- Encourage students to show idea, how much energy is consumed by those appliances.

Stage 3: Explore energy consumption factors in electrical appliance by using “Energy Detective Game”
- Game Case 1: Explore the first factor “Wattage” for energy consumption while time/duration is constrained.
- Game Case 2: Explore the second factor “Time/Duration” for energy consumption while wattage is constrained.
- Check conceptual understanding in identifying the factors of energy consumption in electrical appliances.

Stage 4: Explain the concept of energy consumption
- Summarize the finding from Stage 3;
- Explain the relationship among wattage, voltage, current and time;
- Provide meaning of electrical appliance rating, formula for calculating energy consumption of each appliance beside monthly electric bill.

Stage 5: Elaborate knowledge by applying the constructed knowledge to new situation
- “Energy Efficiency Game”: Monitor your power bill” with different time limitation and scenarios
- Apply the constructed knowledge from Stage 4 to save money on power bill.
- Develop sense of awareness on energy conservation.

Stage 6: Evaluate students’ conceptual understanding
- “Shopping game”
  - Evaluate self-awareness in using electrical efficient appliance.
- Show the game result regarding ways for making appliances have more energy efficient.

Figure 1. Procedure of the game-based conceptual construction with the 5E learning cycle model
Figure 2. Explore energy consumption factors in electrical appliance by using “energy detective game”
Experiment design

Participants

To evaluate the effectiveness of the developed computer game based on the 5E learning cycle model, an experiment was conducted on a secondary school energy consumption and conservation topic. The participants of this experiment were four classes of tenth graders secondary school in eastern Bhutan. A total of 129 students participated in this study. Two classes were assigned to be the experimental group, and the others were the control group. The experimental group included 69 students (31 males and 38 females), while the control group had 60 students (32 males and 28 females). In this study, the same teacher taught the students in the four classes in order to avoid the influence of different experienced teachers on the experimental results. The students in control group learned with the lesson in traditional teaching approach based on 5E learning model supplemented by text books, while those in the experimental group learned with the same lesson supplemented by the developed computer game based on the 5E learning cycle model.

Research tools

To answer the research questions of this study: a pretest, a posttest, a pre-awareness questionnaire, a post-awareness questionnaire, a questionnaire for measuring the satisfaction toward the developed computer game were implemented as the research tools. Both the pretest and the posttest were designed by three experienced teachers. Each test contained 20 multiple-choice items, and one point was scored for each correct answer; therefore, the total score of the tests was 20. The pretest aimed to evaluate the students’ prior knowledge of the energy consumption and conservation content. On the other hand, the posttest aimed to evaluate the learning achievement of the students after participating in the learning activity.

The pre-awareness questionnaire is the same as the post-awareness questionnaire. The awareness questionnaire was adapted from Kang, Cho and Kim (2012). It consisted of fifty items on a 3-point Likert scale in which “3” represents “know very well,” “2” represents “know some of them,” and “1” represents “don’t know at all.” The Cronbach’s alpha value for the awareness questionnaire was 0.99, indicating good reliability in internal consistency.
The questionnaire of satisfaction for the developed educational computer game was used to evaluate the game-based learning with the 5E learning cycle model after the experiment. There were 18 items in three dimensions of the questionnaire: interest (two items), participation (three items), and satisfaction (thirteen items). This questionnaire was measuring using a 5-points Likert scale. The Cronbach’s alpha value for the satisfaction questionnaire was of 0.877, showing good reliability in internal consistency.

**Figure 4.** Evaluate students’ conceptual understanding by using “shopping game”

**Experimental process**

The experiment was conducted on the “energy consumption and conservation” unit of a secondary school physics course, which aimed to teach the students understanding of the main factors of electrical energy consumption in electrical appliances, calculating the energy consumption and cost of operating daily used household electrical appliances, and inculcating sense of awareness on the energy conservation.

Before the experiment, the students took the pretest for evaluating their prior knowledge of the energy consumption and conservation; moreover, a pre-awareness questionnaire was conducted. The learning activities lasted 90 minutes. The learning content for both the experimental and control group was the same. The students in experimental group learned the content supplemented with the developed educational computer game-based on the 5E learning cycle model, whereas those in the control group were taught with conventional teaching method supplemented with text book. After learning activity, a posttest was conducted; moreover, the students were asked to response the post-awareness questionnaire again. In addition, the students of the experimental group completed the satisfaction questionnaire.

**Experimental results**

**Students’ achievement**

Before conducting the inferential statistic tests, we found that pre-test scores from both the control group and the experimental group were normally distributed by Shapiro-Wilk test of $p > 0.05$. Therefore, the use of $t$-test was
appropriate. It was found that the mean ± standard deviation of pretest of the experimental group was 7.52 ± 2.244, and of control group was 7.61 ± 2.289. There was no significant difference between the mean score of pretest of the control and the experimental groups \((t = 0.230, p = 0.409)\), indicating that the students in both the groups had similar prior knowledge/concept regarding energy consumption and conservation.

In order to explore the effect of each treatment, another analysis was made to compare the learning improvement of the students in the two groups (i.e., Control and experimental groups). Before conducting the inferential statistic test, in control group, the normal distribution of pre- and post-test scores was tested. We found that the pretest scores were normally distributed by Shapiro-Wilk test \((p = 0.156)\) but the posttest scores were not normally distributed as indicated by Shapiro-Wilk test \((p = 0.021)\). Similarly, in the experimental group, the pretest scores were normally distributed by Shapiro-Wilk test \((p = 0.148)\), whereas the posttest scores were not normally distributed by Shapiro-Wilk test \((p = 0.002)\). Therefore, we deal with non-parametric hypothesis test, Wilcoxon test is used to analyze the pre- and post-test results as shown in Table 1. The results of the learning improvement (posttest vs pretest) were significant in both the control group and the experimental group. It is clear that the students in the two groups improved their learning achievement after participating in both the traditional learning activity and the developed learning activity.

**Table 1. Wilcoxon result of the pretest vs the posttest**

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Z</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control group</td>
<td>60</td>
<td>7.52 ± 2.244</td>
<td>10.08 ± 3.33</td>
<td>3.775</td>
<td>0.000*</td>
</tr>
<tr>
<td>Experimental group</td>
<td>69</td>
<td>7.61 ± 2.289</td>
<td>11.97 ± 3.83</td>
<td>6.091</td>
<td>0.000*</td>
</tr>
</tbody>
</table>

\* \(p < 0.05\).

Furthermore, to examine how the learning achievement was affected by the treatments after the implementation of the developed learning unit, the posttest scores of both control and experimental groups were analyzed with non-parametric hypothesis test, Mann-Whitney U test, as shown in Table 2.

**Table 2. Mann-Whitney U result of the posttest**

<table>
<thead>
<tr>
<th>Groups</th>
<th>N</th>
<th>Mean ± SD</th>
<th>Z</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control group</td>
<td>52</td>
<td>10.08 ± 3.33</td>
<td>2.537</td>
<td>0.005*</td>
</tr>
<tr>
<td>Experimental group</td>
<td>69</td>
<td>11.97 ± 3.83</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\* \(p < 0.05\).

The result in Table 2 shows that there is significant difference of the posttest scores between the experimental group and the control group. In other words, the mean score of posttest for the experimental group were significantly higher than that for the control group, suggesting that the developed educational computer game based on the 5E learning cycle model could enhance better the learning achievement in energy consumption and conservation unit.

**Energy conservation awareness**

Table 3 shows the descriptive statistics and \(t\)-test of the pre-awareness survey, it shows that students have fairly good ideas about the awareness on energy conservation such as “awareness on heating rooms with minimum energy lost (heating room),” “awareness on using different electrical appliances for cooling rooms (cooling),” “awareness on daily life habits in operating household appliances to minimize energy conservation (household appliances),” but they had comparatively lower awareness about the “awareness on windows insulation with minimum energy(window),” “awareness on lighting rooms for less energy consumption (Lighting),” “Awareness on buying electrical appliances looking into Energy Star Rating for it energy efficiency (Energy Stars).” When classifying this into groups, the results suggest that the students in both the groups had similar level of awareness towards energy conservation. Statistically, there is no significant difference in students’ pre-awareness on energy conservation in both groups except in term of the “cooling” factor, indicating that students in the control group had exhibited better awareness in term of using electrical appliances for cooling rooms than that those in experimental group. This might be because of the students in the experiment group might not have experienced with those appliances that are utilized for cooling the rooms which are especially operated in hot environments.
Table 3. Comparison of students’ pre-awareness on energy conservation

<table>
<thead>
<tr>
<th>Awareness</th>
<th>Mean (SD)</th>
<th>Control Group</th>
<th>Exp. Group</th>
<th>t</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Window: Awareness on windows insulation with minimum energy lost</td>
<td>1.80(0.31)</td>
<td>1.85(0.32)</td>
<td>1.76(0.31)</td>
<td>1.63</td>
<td>0.052</td>
</tr>
<tr>
<td>Room Heating: Awareness on heating rooms with minimum energy lost</td>
<td>2.11(0.44)</td>
<td>2.17(0.43)</td>
<td>2.05(0.44)</td>
<td>1.53</td>
<td>0.065</td>
</tr>
<tr>
<td>Cooling: Awareness on using different electrical appliances for cooling rooms</td>
<td>2.01(0.46)</td>
<td>2.09(0.43)</td>
<td>1.92(0.53)</td>
<td>1.93</td>
<td>0.028*</td>
</tr>
<tr>
<td>Lighting: Awareness on lighting rooms for less energy consumption</td>
<td>1.96(0.58)</td>
<td>1.88(0.58)</td>
<td>2.04(0.57)</td>
<td>1.57</td>
<td>0.060</td>
</tr>
<tr>
<td>Household Appliances: Awareness on daily life habits in operating household appliances to minimize energy conservation</td>
<td>2.05(0.31)</td>
<td>2.08(0.32)</td>
<td>2.02(0.30)</td>
<td>1.05</td>
<td>0.149</td>
</tr>
<tr>
<td>Energy Stars: Awareness on buying electrical appliances looking into Energy Star Rating for it energy efficiency</td>
<td>1.91(0.70)</td>
<td>1.87(0.72)</td>
<td>1.95(0.67)</td>
<td>0.61</td>
<td>0.272</td>
</tr>
</tbody>
</table>

*p < 0.05.

Table 4 shows the students’ level of awareness on energy conservation who participated in two different interventions; it advocates that overall students’ consciousness on electrical energy conservation had enhanced after taking part in this study. The results also reflected that the students in experimental group exhibited better awareness than those in control group. Moreover, it indicates that students in experimental group shows significantly higher level of awareness/consciousness about electric energy conservation, such as in “awareness on daily life habits in operating household appliances to minimize energy conservation (household appliances),” “awareness on buying electrical appliances looking into Energy Star Rating for it energy efficiency (energy stars),” and “awareness on lighting rooms for less energy consumption (lighting)” than those in control group. These are the common factors that contribute majorly in saving energy in our daily life. Statistically, there are significant differences of post-awareness on energy conservation between students in the experimental group and the control group. These results confirms that the developed educational computer game based on the 5E learning cycle model is able to promote students’ awareness on electric energy conservation, particularly in buying and using daily household appliances in their life as to save money in long time.

Table 4. Comparison of students’ post-awareness on energy conservation

<table>
<thead>
<tr>
<th>Awareness</th>
<th>Mean (SD)</th>
<th>Control Group</th>
<th>Exp. Group</th>
<th>t</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Window: Awareness on windows insulation with minimum energy lost</td>
<td>2.55(0.40)</td>
<td>2.48(0.44)</td>
<td>2.62(0.37)</td>
<td>-1.97</td>
<td>0.03*</td>
</tr>
<tr>
<td>Room Heating: Awareness on heating rooms with minimum energy lost</td>
<td>2.43(0.40)</td>
<td>2.36(0.42)</td>
<td>2.50(0.39)</td>
<td>-1.95</td>
<td>0.03*</td>
</tr>
<tr>
<td>Cooling: Awareness on using different electrical appliances for cooling rooms</td>
<td>2.50(0.47)</td>
<td>2.41(0.58)</td>
<td>2.60(0.36)</td>
<td>-2.19</td>
<td>0.02*</td>
</tr>
<tr>
<td>Lighting: Awareness on lighting rooms for less energy consumption.</td>
<td>2.66(0.46)</td>
<td>2.56(0.58)</td>
<td>2.75(0.34)</td>
<td>-2.38</td>
<td>0.01*</td>
</tr>
<tr>
<td>Household Appliances: Awareness on daily life habits in operating household appliances to minimize energy conservation.</td>
<td>2.56(0.31)</td>
<td>2.53(0.36)</td>
<td>2.60(0.26)</td>
<td>-2.93</td>
<td>0.00*</td>
</tr>
<tr>
<td>Energy Stars: Awareness on buying electrical appliances looking into Energy Star Rating for it energy efficiency.</td>
<td>2.47(0.51)</td>
<td>2.40(0.62)</td>
<td>2.54(0.41)</td>
<td>-2.51</td>
<td>0.01*</td>
</tr>
</tbody>
</table>

*p < 0.05.
Satisfaction with the educational computer game

Table 5 shows the descriptive statistics of the satisfaction questionnaire. Overall analysis from the questionnaires reveals that the students have rated “high satisfaction” for the developed computer game-based learning for being able to develop interest in learning the concept of electrical energy consumption and conservation. Moreover, the students stated that they were “highest satisfaction” towards the developed computer game based on the 5E learning cycle model as it had encouraged them to participate in the learning activities through its interaction; in addition, the students rated that they have “high” level of satisfaction in learning concept of electrical energy consumption and conservation through the developed educational computer game.

Table 5. Students’ satisfaction about the developed educational computer game

<table>
<thead>
<tr>
<th>Satisfaction</th>
<th>Mean</th>
<th>SD</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Interest</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I enjoy physic lessons very much with computer game.</td>
<td>4.27</td>
<td>0.78</td>
<td>High satisfaction</td>
</tr>
<tr>
<td>I become more curious and observant in the class when the lesson is integrated with computer game.</td>
<td>4.39</td>
<td>0.76</td>
<td>High satisfaction</td>
</tr>
<tr>
<td><strong>Participation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I enjoy participating in class activities when the lessons are taught using computer game.</td>
<td>4.45</td>
<td>0.77</td>
<td>High satisfaction</td>
</tr>
<tr>
<td>The lessons using animations, game and simulated activities make me more attentive in the class.</td>
<td>4.56</td>
<td>0.73</td>
<td>Highest satisfaction</td>
</tr>
<tr>
<td>Integration of computer game in the lesson promotes better interaction amongst friends and teachers.</td>
<td>4.33</td>
<td>0.73</td>
<td>High satisfaction</td>
</tr>
<tr>
<td><strong>Satisfaction</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>It is easier for me to understand the content with computer game.</td>
<td>4.45</td>
<td>0.59</td>
<td>High satisfaction</td>
</tr>
<tr>
<td>I get learning satisfaction when I learn the lesson with computer game.</td>
<td>4.38</td>
<td>0.67</td>
<td>High satisfaction</td>
</tr>
<tr>
<td>Computer game in the lesson helps me to develop confidence in learning electrical energy calculation.</td>
<td>4.53</td>
<td>0.75</td>
<td>Highest satisfaction</td>
</tr>
<tr>
<td>I found the graphics and computer game useful in visualizing the concepts.</td>
<td>4.26</td>
<td>0.69</td>
<td>High satisfaction</td>
</tr>
<tr>
<td>Computer game in learning helps me to think and analyze the real things in world.</td>
<td>4.48</td>
<td>0.71</td>
<td>High satisfaction</td>
</tr>
<tr>
<td>It helps me to develop the relevance between the course and real world situations.</td>
<td>4.20</td>
<td>0.85</td>
<td>High satisfaction</td>
</tr>
<tr>
<td>Computer game allows me to develop skills needed in the real world.</td>
<td>4.39</td>
<td>0.72</td>
<td>High satisfaction</td>
</tr>
<tr>
<td>I like the way the teacher uses computer game to teach energy consumption by various household appliance lesson.</td>
<td>4.70</td>
<td>0.58</td>
<td>Highest satisfaction</td>
</tr>
<tr>
<td>The use of computer game in the lesson helps me to build confidence in understanding the concept of energy consumption clearly.</td>
<td>4.32</td>
<td>0.68</td>
<td>Highest satisfaction</td>
</tr>
<tr>
<td>I like computer game in learning electrical energy consumption and conservation sessions because it enables me to learn faster.</td>
<td>4.53</td>
<td>0.71</td>
<td>High satisfaction</td>
</tr>
<tr>
<td>I gain confidence when I learn the lesson using computer game.</td>
<td>4.23</td>
<td>0.65</td>
<td>High satisfaction</td>
</tr>
<tr>
<td>I like electrical energy lessons with computer game because the lessons are interesting, informative and help to visualize the abstract concepts of energy better.</td>
<td>4.48</td>
<td>0.71</td>
<td>High satisfaction</td>
</tr>
<tr>
<td>I like computer game technology integration in all the subjects to help enhance our critical thinking.</td>
<td>4.30</td>
<td>0.84</td>
<td>High satisfaction</td>
</tr>
</tbody>
</table>
Discussion

This study was conducted to examine the effectiveness of two methods of learning electric energy consumption and conservation (computer game based on a learning cycle approach and conventional teaching approach) for secondary school students. The results of this study revealed that the learning cycle-based computer game group significantly outperformed the conventional group in understanding main factors of energy consumption in electrical appliances, calculating the energy consumption and cost of operating daily used household electrical appliances, and inculcating sense of awareness on the energy conservation. Possible reasons for this observed difference might include the value related with alternative ways of acquiring and constructing knowledge in the topic, particularly exploration, and elaboration. During the educational game, students learned through their own actions and reactions by being involved in gamified activities. They explored the energy consumption factors in electrical appliances and the graphical responses from simulating the energy consumption encouraged them to identify the correct factors. Students’ explorations involved trying out and learning from errors for saving money on power bill as much as they can. They were also involved in gamified activities that helped them to examine the adequacy of their conceptions and encouraged them to discuss about those conceptions. This led to the opportunity to construct more appropriate concepts and develop energy conservation awareness. Meanwhile, students in the conventional teaching group mainly focused on concepts related to the topic, the process that required less conceptual restructuring.

The findings of this study showed that the developed educational computer game provided higher level of learning achievement, awareness and satisfaction for students. This was consistent with various studies that advocated the correct use of the learning cycle instruction accomplished both effective learning of concepts and an ability to apply concepts (Piyayodilokchai et al., 2013; Yadigaroglu & Demircioglu, 2012) and with the study that used a game-based learning strategy to encourage learners to conserve home energy usage (Yang et al, 2012). Moreover, during involving in gamified activities, they become aware of their own ideas, beliefs and attitudes by using the simple tasks which is related to everyday tasks or phenomena (Ebenezer, Chacko, Kaya, Koya, & Ebenezer, 2010). Therefore, the students are engaged and motivated via direct experiences. In other words, it was also suggested that developing usable educational technology showed a more seamless learning experience (Slotta & Najafi, 2012).

However, there is no report of the study that applies the 5E learning cycle model in developing an educational computer game. So this study may provide a useful guideline to further the development of a computer game based on the 5E learning cycle model.

Conclusion

The result of this study indicated that the use of educational computer game embedded with the 5E learning cycle model is an alternative teaching and learning method to traditional instruction and could be used to promote meaningful learning in electric energy consumption and conservation. From the effective learning outcomes of this study, we recommend a set of guidelines for any practitioner/teacher/instructor as follows:

- The lesson is introduced with the scenario-based, so teacher(s) should emphasis on the objective of the scene with further explanation to develop curiosity for learning the lesson (e.g. “Why monthly electric bill is more?”). This should not take more than 5 minutes.

- Participants should be instructed to follow the learning modules in sequence as Energy Detective, Energy Efficiency Game, and Shopping Game. The time allocation for these activities is assigned for 50 minutes tentatively as to suit the time needed by the students for different activities depending upon the ability of the students.

- At the end of each phase, the teacher(s) need to provide participants opportunities to share and explain their understanding of the concepts, skills and processes they are learning. In the meantime, teacher(s) should take opportunities to clarify students’ misconceptions and help them in introducing new concepts, process and skills. Moreover, teacher(s) should challenge and extends students’ conceptual understanding and skills of the concept by conducting additional activities that focuses on adding breadth and depth to current understanding.
The learning process is a vital aspect of the developed supplemented instructional unit. Therefore, designing a formative assessment that suits the students’ background and learning environment to follow students learning process to help in evaluating more authentic learning outcomes are required.

Beside the mentioned points, it is always advisable that teachers/practitioners inform the students about the learning process of the instructional unit before the implementations, as the new type of teaching and learning approach might not be familiar for some students in some schools. This would minimize confusion and time wastage, and eventually help in setting good flow while implementing the learning unit. Still if any teachers/instructors/practitioners having difficulty in executing the teaching and learning activities related to gamified intervention embedded with 5E learning cycle model, the authors will be willing to provide a guided lesson plans, a manual of the game or technical assistance upon request. Moreover, to be generalization, it should as well be studied more on a large number of students with different backgrounds before being adopted for wider use.

Acknowledgements

This research project is supported by Thailand International Cooperation and Development Agency (TICA) scholarship.

References


The Impact of Digital Mobile Devices in Higher Education

M.ª Luisa Sevillano-García and Esteban Vázquez-Cano*
Department of Didactics and School Organization, Faculty of Education, Spanish National University of Distance Education (UNED), C/Juan del Rosal, 14 (28040) Madrid, Spain // mlsevillano@edu.uned.es // evazquez@edu.uned.es

*Corresponding author

(Submitted January 7, 2014; Revised May 6, 2014; Accepted July 1, 2014)

ABSTRACT
This research examined the acceptance, incidence, and use of digital mobile devices (tablets and smartphones) among university students in the European Higher Education Area (EHEA). The research was contextualized in a sample of 419 students from three Spanish public universities. Through a quantitative methodology, we identified the factors and variables that condition and favor the use of digital mobile devices in university studies. The factor-based structure enabled a holistic view of ICT innovation in universities across three key areas: Teaching model (face-to-face/distance), area of study, and generic competencies. The outcomes could be used as means to forecast, explain, and improve the integration of these digital mobile devices for fostering learning activities and generic competencies in the Higher Education context.

Key words
Digital mobile devices, Mobility, Ubiquity, Generic competencies, EHEA

Introduction

This paper analyzes internal and external variables that condition the use, functionality, and acceptance of digital mobile devices (DMD: tablets and smartphones) as tools and resources potentially beneficial in the development of university studies within the European Higher Education Area (EHEA). For this purpose, we have used an ad hoc questionnaire to analyze the incidence and factors that condition the use of DMD in three Spanish universities: University Complutense of Madrid, University of Oviedo, and Spanish National University of Distance Education (UNED). Digital mobile devices have burst in Spain, a country in Europe, with the highest penetration of tablets and smartphones (66%), 9 points higher than the European average (57%) (Mobile Life Report, 2013). The use of DMD applied to the development of personal, academic, and professional competencies is part of EHEA’s philosophy, which should promote in students the acquisition of a series of transversal or generic skills that secure an appropriate level of performance in terms of academic and professional competencies according to current societal demands. The key concept adopted in this framework is the use of university methodologies, tools, and resources that encourage the transferability of skills to personal, social, academic, and professional contexts and, thus, being able to create an effective foundation for life-long learning (Bennett, & Maton, 2010; Cedefop, 2010; Allen, & van der Velden, 2012).

All experts suggest that the right way to overcome this economic crisis is to intelligently adopt digital technologies to develop productive and efficient economic models where technological innovation serves to promote growth and productivity (Networked Society Report, 2013). Digital mobile devices have been evolving at a swift pace, and educational society needs time to assimilate and evaluate DMD possibilities in education (Weider, 2011; Gawelek, Komarny, & Spataro, 2011; Vázquez-Cano, 2012, 2014). Currently, DMD applications are multiplying, and their continual development opens up many possibilities in the field of education, encouraging socialization and inclusive actions for a great variety of groups. Nowadays, citizens’ training requires specific attention to foster the knowledge necessary to make decisions on the use of objects and technological processes, solve problems, and use them to increase the capacity of knowing how to act and use them to achieve a better learning (Bertchy, Cattaneo, & Wolter, 2009; Murphy, 2011; Rasmus, 2013). Therefore, our objective in this research is to determine variables and factors that affect the academic use of digital mobile devices (Tablets and Smartphones) by students in EHEA.

Digital mobile devices in the European higher education area

For some years now, there has been intense scientific activity focusing on the impact of Information and Communications Technology on Education and the factors that facilitate or impede said impact (Balanskat, Blamire,
These studies all agree that ICT tools and resources may have a significant influence on teaching-learning processes. This context, mediated by ICT, promotes that knowledge stops to be associated to specific physical spaces and persons and goes through concepts of “mobility” and “ubiquity” (Cope, & Kalantzis, 2009; Mercier, & Higgins, 2013). This enables new scenarios that relax the pace of work, training facilities, and learning interests. On the other hand, it also enables the student to become a digital content-creator, to develop his/her conception of knowledge within an area of personal learning, and spread it (Van't, & Swan, 2007; Thorpe, & Gordon, 2012). This trend is reflected in recent international research, as the one performed in 2012 by the ARD/ZDF (Germany) that concludes that mobile terminals, especially “Tablets” and “Smartphones” experienced an important development in Germany, increasing and making easier access to multiple contents on networks (Eimeren, & Frees, 2012).

What is clear is that educational innovation through ICT will not only be achieved through administrative orders, regardless of whether those orders are passed down from the European, national or regional authorities. It is professors themselves who can and must adapt and incorporate these initiatives into the development of their university subjects. Teachers must renew their methodological training and be able to guide students toward methodologies that promote autonomous interaction and collaborative work. Though viewing each technological breakthrough as the quintessence of educational innovation is a mistake; other variables must also be considered, such as existing context and culture, educational practices and relations, and the way those involved conceive of teaching-learning processes. This restates the dynamics of educational activities, giving special value to independent and non-classroom activities (Sevillano, & Quicios, 2012). Learning can be restructured and adapted from the principle of ubiquity, but for this challenge, institutions need to orientate methodologies toward the use of new mobile devices, from the possibilities offered primarily through open educational resources (OERs) distributed on wikis, blogs, mash-ups, podcasts, social software, virtual worlds, personal learning environments (PLEs), massive open online courses (MOOCs), and other emerging online practices. In this digital context, technologies offer a new media format that supports new ways of narrating contents (Katz, 2013; Vázquez-Cano, López-Meneses, & Sarasola, 2013). From the students’ perspective, the key factors for innovation in Education will be the need for competition and implementing the mobility needs for a global higher education without frontiers as it is being promoted today with the trend of MOOCs and PLEs. Mobile computing environments appear to be more student-centered (Norris, & Soloway, 2004). Research shows that constructivist teaching practices are more prevalent in these contexts (Cambre, & Hawkes, 2004; UNESCO, 2013). Attendance rates improve and disciplinary referrals decline (Knezek, & Christensen, 2005). Student attitudes toward learning also improve (UNESCO, 2013) and the use of project-based and inquiry-based lessons increase with the use of ubiquitous devices (Norris, & Soloway, 2004). Ubiquitous learning (Barbosa, Barbosa, & Wagner, 2012) represents a new educational paradigm that, to a large extent, is made possible by new media and digital instruments. Ubiquity and mobility become recurrent principles for educational performance in this century. There is a direct relation between the idea of ubiquitous learning and the ability of mobile devices to provide highly interconnected educational environments (Hwang, Kuo, Yin, & Chuang, 2010; Chia-Ching, & Chin-Chung, 2012; Johnson et al., 2013).

In this educational and digital scenario, the European Union emphasizes that each citizen will need a wide range of competencies to adapt with flexibility to a world that is rapidly changing and is highly interconnected (European Commission, 2010). A student who performs a practice in a ubiquitous digital ecosystem is subjected to a series of stimuli flowing between nodes located in different media that filter information through a variety of channels, each one with individual narratives and symbolic codes (Cheung, & Vogel, 2013; Ahmed, & Parsons, 2013). In this ecology of predominate user-generated contexts (Wunsch-Vincent, & Vickery, 2007), users are actively involved in developing their own forms of content (Gegenfurtner, Veermans, & Vauras, 2013). The use of DMD, mainly tablets and smartphones, on learning processes leads to situations where ubiquitous devices and training activities in applied contexts of social practice converge. We have reflected in Table 1 the relation established between applications and most common uses of DMD and their relation to EHEA's general competencies (Dublin Descriptors, 2004).

<table>
<thead>
<tr>
<th>DMD Capabilities (Tablets and Smartphones)</th>
<th>Generic Competencies (Dublin Descriptors, 2004)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Management of social networks (Twitter, Facebook, LinkedIn ...) and all kinds of forums.</td>
<td>Self-regulated learning competence</td>
</tr>
<tr>
<td>Videoconference (especially easy on iPad devices with FaceTime application), sending short messages and phone calls over the Internet (VoIP) without additional costs.</td>
<td>Self study.</td>
</tr>
<tr>
<td>Planning and organization.</td>
<td>Initiative and motivation.</td>
</tr>
<tr>
<td>Adequate management of time.</td>
<td></td>
</tr>
<tr>
<td>Concern for quality (monitoring, evaluation, improvement).</td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Reading books and digital documents using office applications, e-readers, microblogging, readers of RSS subscriptions.</td>
<td></td>
</tr>
<tr>
<td>Using the GPS locator, map viewer (Google Maps, Google Earth ...), trace-router.</td>
<td></td>
</tr>
<tr>
<td>Environment Sensors: weather station, gyroscope, and accelerometer, compass and magnetic field sensing, light level measurement and noise.</td>
<td></td>
</tr>
<tr>
<td>Augmented reality applications.</td>
<td></td>
</tr>
<tr>
<td>Higher cognitive Competencies</td>
<td></td>
</tr>
<tr>
<td>Ability to summarize.</td>
<td></td>
</tr>
<tr>
<td>Ability to analyze.</td>
<td></td>
</tr>
<tr>
<td>Creative and innovative thinking.</td>
<td></td>
</tr>
<tr>
<td>Ability to solve problems / act in new environments.</td>
<td></td>
</tr>
<tr>
<td>Application of knowledge to practice.</td>
<td></td>
</tr>
<tr>
<td>Decision-making.</td>
<td></td>
</tr>
<tr>
<td>Critical judgment (personal work and the work of others).</td>
<td></td>
</tr>
<tr>
<td>Communication Competencies</td>
<td></td>
</tr>
<tr>
<td>Videoconferences. Play and create FM radio: images, videos, animations, music and other audio files.</td>
<td></td>
</tr>
<tr>
<td>Sound recording and photo editing, video and sound editing (with various apps).</td>
<td></td>
</tr>
<tr>
<td>Translation of words and texts in multiple languages (with voice playback and possibility of audio input words).</td>
<td></td>
</tr>
<tr>
<td>Publications tools (Calameo, Scoop it, etc.).</td>
<td></td>
</tr>
<tr>
<td>Augmented reality applications.</td>
<td></td>
</tr>
<tr>
<td>Instrumental Competencies in Knowledge Society</td>
<td></td>
</tr>
<tr>
<td>ICT Management.</td>
<td></td>
</tr>
<tr>
<td>Ability to find information.</td>
<td></td>
</tr>
<tr>
<td>Skill in organizing information.</td>
<td></td>
</tr>
<tr>
<td>Ability to manage databases.</td>
<td></td>
</tr>
<tr>
<td>Interpersonal Competencies</td>
<td></td>
</tr>
<tr>
<td>Management (sync “in the cloud”) e-mail accounts, contacts and calendar / on-line user agenda.</td>
<td></td>
</tr>
<tr>
<td>Management of social networks (Twitter, Facebook, LinkedIn ...) and all kinds of forums and chats.</td>
<td></td>
</tr>
<tr>
<td>Videoconference.</td>
<td></td>
</tr>
<tr>
<td>Use other tools to develop documents and share blogs, posters, concept maps, comics, and stories.</td>
<td></td>
</tr>
<tr>
<td>Interpersonal Competencies</td>
<td></td>
</tr>
<tr>
<td>Ability to negotiate effectively / conflict resolution.</td>
<td></td>
</tr>
<tr>
<td>Ability to coordinate.</td>
<td></td>
</tr>
<tr>
<td>Ability to work as a team / collaborative work.</td>
<td></td>
</tr>
</tbody>
</table>

While the previous discussion shows positive effects of ICT in Higher Education, the benefits of using DMD are sporadic and even inconsistent (Moran, Hawles, & Gayar, 2010). The lack of quality professional development, university policies that fail to support use of mobile technology, instructors’ beliefs about the role of technology in subject development, and cultures that are just not supportive of mobile computing adoption have made the integration of mobile initiatives problematic in Spanish universities. As a result, researchers have pursued better technology acceptance models capable of delivering higher prediction successes (Legris, Ingham, & Collerette, 2003). Refinements will be necessary to overcome barriers that might prevent an individual from adopting a particular information system technology, such as system design characteristics, training, support, and decision-maker characteristics (Baird, & Fisher, 2006; EACEA/Eurydice, 2011,2012). Examination of other literature on technology acceptance indicates that social variables such as demographics, managerial knowledge, environmental characteristics, and task-related characteristics would also expand the model’s predictive capabilities (Pijpers, 2001; Pachler, Bachmair, & Cook, 2010). Research on the effects of DMD on students has been described as the most mature research area in contemporary information systems research literature (BECTA, 2009; Moran, Hawles, & Gayar, 2010; Goral, 2011; Eichenlaub et al., 2011; Yoiro, & Feifei, 2012).

Thus, we have developed a specific questionnaire within this framework to measure the following two objectives:

- To identify the factors that foster innovation with DMD among students at university, through the development of a scale and the subsequent analysis of its psychometric properties, such as validity and reliability.
- To analyze the DMDs’ incidence on developing learning activities and generic competencies according to EHEA principles.
**Study context**

The population of this study comprised 419 higher education students from Graduate and Master studies at three Spanish universities (*University Complutense of Madrid, University of Oviedo and National University of Distance Education-UNED*). University Complutense of Madrid and Spanish National University of Distance Education (UNED) are the two biggest universities in Spain. Both have more than 500,000 students and UNED is the biggest distance university in Europe with more than 260,000 students around the world. The university degrees selected were: English and Spanish Philology, Geography and Art, Architecture, Industrial Engineering, Pedagogy, Social education, Biology and Nursing.

To identify students using digital mobile devices, a general digital questionnaire was previously conducted in these three universities. The students were selected based on availability and willingness to participate. When the survey was launched, the invitations were sent by email to all registered students and they were asked to participate on a voluntary basis without any incentives. A total of 956 questionnaires were received and based on this incidental sample, a list of 419 students who possessed DMD was drawn up, thus, clearly identifying the target population of this study. The sample is given in Table 2.

**Table 2. Descriptive Statistics of the Sample in Terms of Frequencies and Percentages**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Total</th>
<th>Educational Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Freq.</td>
<td>%</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>201</td>
<td>47.97</td>
</tr>
<tr>
<td>Women</td>
<td>218</td>
<td>52.02</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-20</td>
<td>131</td>
<td>31.26</td>
</tr>
<tr>
<td>21-23</td>
<td>145</td>
<td>34.60</td>
</tr>
<tr>
<td>23-25</td>
<td>101</td>
<td>24.10</td>
</tr>
<tr>
<td>≥ 26</td>
<td>42</td>
<td>10.02</td>
</tr>
<tr>
<td>Studies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Science</td>
<td>139</td>
<td>33.17</td>
</tr>
<tr>
<td>Social Sciences</td>
<td>181</td>
<td>43.19</td>
</tr>
<tr>
<td>Humanities</td>
<td>99</td>
<td>23.62</td>
</tr>
<tr>
<td>Total</td>
<td>419</td>
<td>100</td>
</tr>
</tbody>
</table>

**Research design**

The objective of this study was to measure factors that condition the use of DMD by university students in EHEA. For this purpose, we developed and implemented an ad hoc questionnaire called: Scale of Factors that Foster Innovation with DMD (SFFIDMD) among university students during this research as funded by the Spanish Ministry of Education (EDU2010-17420-Sub EDUC). Research in this area has generated adoption metrics that can be used to determine the probability of successful implementation of information system initiatives. Gathering of views, opinions, and attitudes toward ICT by means of “ad hoc” questionnaires (Balanskat, Blamire, & Kefala, 2006; Drent, & Meelissen, 2008) is one of the strategies used for identifying factors that facilitate ICT implementation, use, and innovation. Several previous studies have shown that there are various external factors that influence the adoption of innovation (Venkatesh et al. 2003; Kearney et al., 2012; Voogt et al. 2013). In this study we expect DMDs could be an important factor for developing digital didactic strategies for higher education. For this purpose, we have proposed a model (Figure 1) with the support of LISREL software which makes possible to estimate many parameters in complex structures of interaction.

Therefore, apart from a series of descriptive general and academic data, such as age and sex, area of study, and type of university (face-to-face or distance), the questionnaire comprised 29 items, to which participants responded using a five-point Likert-type assessment scale. The five response options are as follows: 1 = I totally disagree; 2 = I disagree; 3 = I neither agree nor disagree; 4 = I agree; and 5 = I totally agree. The students completing the questionnaire were asked to indicate the extent to which they agreed or disagree with the different statements.

The data provided in questionnaire were analyzed using version 19 of the SPSS statistical software package. The questionnaire’s statistical guarantees were studied. The item-total correlation of the dimension was analyzed to
eliminate those items with a correlation coefficient of below 0.2. In addition, the reliability of the scale was analyzed using the Cronbach’s Alpha Test (.843). Next, a factor analysis of principal components was conducted to determine the internal structure of the questionnaire. However, before carrying out the analysis, and as a prior statistical requisite that guarantees correct application, a series of other tests were performed: First, Bartlett’s test of sphericity, was used to test the hypothesis that the correlation matrix obtained is not an identity matrix; in other words, that significant interrelations exist between items that justify the factor analysis. Secondly, the KMO (Kaiser-Mayer-Olkin) index, which measures sampling adequacy (data suitability) to carry out the factor analysis, was used. Finally, the reliability of the factors extracted from the questionnaire was analyzed, both individually and together. In addition, an inferential analysis was conducted by comparing the means of the questionnaire factors in relation to sex, age, area of study (Humanities, Science, and Social Sciences), and main teaching model (face-to-face or distance). To this end, the Student’s T-test was used in all the variables for independent samples, with the exception of area of study in which a single-factor analysis of variance (ANOVA) was used.

The following three hypotheses are proposed:

H1. DMD have a positive effect on the development of generic competencies.
H2. DMD have a positive effect on developing learning activities.
H3. Performance, effort expectancy, and self-efficacy are high when using DMD in Higher Education.

Results

The results obtained in the statistical tests applied to the questionnaire attest to its internal consistency and construct validity. None of the items were eliminated as a result of low discriminatory power or low correlation with the dimension as a whole. Bartlett’s test of sphericity ($p = .000$) and the Kaiser-Mayer-Olkin sampling adequacy measure (.767) were found to be suitable when analyzing the factorial structure of the scale using the Varimix with Kaiser Normalization method for the principal component analysis.

<table>
<thead>
<tr>
<th>Item</th>
<th>PE1. Enable me to accomplish tasks more quickly.</th>
<th>PE2. Hamper my performance and effectiveness in class.</th>
<th>PE3. Because my classmates perceive me as competent.</th>
<th>PE4. Increase the instructors respect for me.</th>
<th>PE5. Decrease my chances of getting a good grade.</th>
<th>EE-A1. Learning to operate the DMD is easy for me.</th>
<th>F1 (PE)</th>
<th>F2 (EE)</th>
<th>F3 (SE)</th>
<th>F4 (COM)</th>
<th>F5 (LA)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.854</td>
<td>.432</td>
<td>.501</td>
<td>.534</td>
<td>.401</td>
<td>.790</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. Proposed structure model

Table 3. Matrix of Factors Extracted by Varimax Rotation and Factor Loadings of Items
EE-A2. I find the DMD to be flexible to interact with. .801
EE-A3. Using DMD take too much time from my normal duties. .503
EE-A4. Using DMD is a good idea. .678
EE-A5. I dislike the idea of using DMD. .412
SE 1. I have the resources necessary to use DMD. .734
SE 2. I have the knowledge necessary to use DMD. .765
SE 3. Using DMD fits into my work style. .856
SE 4. If I had seen someone else demonstrate how it could be used. .521
SE 5. If I had a lot of time to complete the job. .479
COM 1. Analysis-Synthesis. .825
COM 2. Communicative. .823
COM 3. Digital. .886
COM 4. Foreign Language. .715
COM 5. Mechanicals. .734
COM 7. Technical. .676
COM 8. None. .303
LA 1. Academic Activities. .799
LA 2. Teacher’s Explanations. .753
LA 3. Didactic Materials. .879
LA 4. Resolution of doubts. .741
LA 5. Others. .356
LA 6. None. .289
% of variance explained = 75.44
6.01 8.03 12.34 25.61 23.45
Cronbach’s alpha α .811 .703 .742 .801 .772

Note. Performance Expectancy (PE); Effort Expectancy (EE); Self-Efficacy (SE); Learning activities (LA); Competencies (COM).

The principal component factor analysis identified five underlying factors in the questionnaire, with a total explained variance of 75.44%. Each of the factors identified is described below.

- **Factor 1. Performance Expectancy.** The items of this factor, as they are formulated, tell us about students' reasons and expectations about the use of DMD use at university. A high level of reliability was obtained in item 1: “Accomplish task more quickly” (.854***).

- **Factor 2. Effort Expectancy:** The five items of this factor reflect certain ways of looking at ubiquitous and mobile technology and its impact on learning practices. DMD are viewed as “devices easy to operate” (.790***) and “easy to interact with” (.801***) and what means a clear correlation (.556) with “performance expectancy”. The subscale that assesses this factor has a Cronbach’s alpha = .703.

- **Factor 3. Self-Efficacy.** Items 1 and 3 cover the content of the category because most of students have “Knowledge necessary to use DMD” (.765***) and its use “fit into their work style” (.856***) and “easy to interact with” (.801***). The educational community is very receptive to DMD implementation, ensuring that students who use and innovate with technology enjoy a greater degree of social acceptance. The seven items make up this subscale have a reliability level of alpha = .742.

- **Factor 4. Competencies.** This factor analyzes the generic competencies related to the use of DMD. In this category, four indicators of generic competencies are especially remarkable: “information processing” (.834***), “analysis and synthesis” (.825***), “foreign language” (.715***), and digital (.886***). The development of these indicators of generic competencies through these technological resources is in line with the recommendations made in European projects, such as Tuning, Reflex, and UE-Converge, and recent international research (Goral, 2011; Eichenlaub et al., 2011, UNESCO, 2013). These four competencies help to provide an answer to the levels required by the Dublin Descriptors (2004): self-management and self-regulated work, management of information and communication processes, and teamwork developing different types or roles. The seven items (competencies) of this factor together make up one of the most numerous subscales that explain 25.78% of the total variance. Its factor loadings are high, as is its reliability (Cronbach’s alpha = .801).
• Factor 5. Learning Activities. The four items that make up this subscale refer to the most common and beneficial learning activities with the use of DMD. It can be highlighted access to “learning and didactic materials” (.897***) and “academic activities” (.799**). “The use of these digital devices supports academic activities of teachers and students in face-to-face and distance education.” This subscale has a mid-level reliability index (Cronbach’s alpha = .772).

Pearson’s correlations tests were also carried out between the factors that make up the questionnaire. The data obtained from the means of the items of each factor provide information regarding the construct validity and indicate that “Self-Efficacy” and “Effort expectancy” (.053) contribute significantly to measuring the dimensions of the construct on acceptance and incidence of DMD as innovation with ICT at universities. In short, positive and significant (p < .01) correlations were observed between the two factors and the development of “Competencies” and “Learning activities”. As shown in Table 4, the factors that most influence acceptance and incidence in using DMD are “Competencies” and “Learning activities”, with means of 3.43 and 3.28, respectively. On the other hand, the factors that have the least influence are “Performance” and “Effort expectancy”, with means of 2.74 and 2.8, respectively, but with a high correlation. In Tables 4 and 5, we present the differences between means for factors with incidence in accordance with teaching model and area of study.

### Table 4. Differences between the means for factors (Face-to-face and Distance)

<table>
<thead>
<tr>
<th></th>
<th>F1 PE</th>
<th>F2 EE</th>
<th>F3 SE</th>
<th>F4 COM</th>
<th>F5 LA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>2.81</td>
<td>2.73</td>
<td>3.13</td>
<td>3.43</td>
<td>3.28</td>
</tr>
<tr>
<td>N = 419</td>
<td>Sd .98858</td>
<td>.89593</td>
<td>.80681</td>
<td>.68131</td>
<td>.70091</td>
</tr>
<tr>
<td>Face to Face</td>
<td>Mean 2.51</td>
<td>2.57</td>
<td>2.98</td>
<td>3.22</td>
<td>3.10</td>
</tr>
<tr>
<td>N = 219</td>
<td>Sd .97651</td>
<td>.81622</td>
<td>.7999</td>
<td>.73589</td>
<td>.70391</td>
</tr>
<tr>
<td>Distance</td>
<td>Mean 3.02</td>
<td>2.87</td>
<td>3.25</td>
<td>3.59</td>
<td>3.42</td>
</tr>
<tr>
<td>N = 200</td>
<td>Sd .80441</td>
<td>.93518</td>
<td>.79713</td>
<td>.58897</td>
<td>.66805</td>
</tr>
<tr>
<td>Sig.</td>
<td>*</td>
<td>*</td>
<td>**</td>
<td>***</td>
<td>***</td>
</tr>
</tbody>
</table>

**Note.** *p < .05, **p < .01, ***p < .001. Performance Expectancy (PE); Effort Expectancy (EE); Self-Efficacy (SE); Learning activities (LA); Competencies (COM).

No statistically significant differences were observed in relation to participants’ sex. The results (Table 4) do indicate significant differences in relation to the type of teaching, with students in Distance Studies scoring higher in the mean for all factors and especially in factor “Performance Expectancy,” especially the item: “Enable me to accomplish tasks more quickly” (t = −3.621; p = .004). “Effort Expectancy” with the item: “I find the DMD to be flexible to interact with” (t = −3.174; p = .001). “Self-Efficacy” with the item: “Using DMD fits into my work style” (t = −3.812; p = .000). “Learning activities” where “Didactic Materials” is the highest loaded (t = −2.756; p = .017) and “Competencies” with “Communicative” item (t = −2.489; p = .024).

### Table 5. Differences between the means for factors (Science, Social Sciences, and Humanities)

<table>
<thead>
<tr>
<th></th>
<th>F1 PE</th>
<th>F2 EE</th>
<th>F3 SE</th>
<th>F4 COM</th>
<th>F5 LA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>2.81</td>
<td>2.73</td>
<td>3.13</td>
<td>3.43</td>
<td>3.28</td>
</tr>
<tr>
<td>N = 419</td>
<td>Sd .98858</td>
<td>.89593</td>
<td>.80681</td>
<td>.68131</td>
<td>.70091</td>
</tr>
<tr>
<td>Science</td>
<td>Mean 2.98</td>
<td>2.97</td>
<td>3.21</td>
<td>3.52</td>
<td>3.31</td>
</tr>
<tr>
<td>N = 139</td>
<td>Sd .95431</td>
<td>.82100</td>
<td>.80987</td>
<td>.76789</td>
<td>.72361</td>
</tr>
<tr>
<td>Social Sciences</td>
<td>Mean 2.84</td>
<td>2.87</td>
<td>3.19</td>
<td>3.12</td>
<td>3.12</td>
</tr>
<tr>
<td>N = 181</td>
<td>Sd .81324</td>
<td>.92134</td>
<td>.77689</td>
<td>.59800</td>
<td>.67890</td>
</tr>
<tr>
<td>Humanities</td>
<td>Mean 2.02</td>
<td>2.69</td>
<td>2.95</td>
<td>2.89</td>
<td>2.92</td>
</tr>
<tr>
<td>N = 99</td>
<td>Sd .89876</td>
<td>.92345</td>
<td>.78907</td>
<td>.50134</td>
<td>.60134</td>
</tr>
<tr>
<td>Sig.</td>
<td>*</td>
<td>*</td>
<td>**</td>
<td>***</td>
<td>***</td>
</tr>
</tbody>
</table>

**Note.** *p < .05, **p < .01, ***p < .001. Performance Expectancy (PE); Effort Expectancy (EE); Self-Efficacy (SE); Learning activities (LA); Competencies (COM).
The means obtained for each area of study (Table 5) reveal differences between the factors facilitating educational innovation with DMD. An analysis of variance (ANOVA) identified those factors with statistically significant differences of over alpha = .05 in accordance with area of study. In this sense, the scores for the factor “Competencies” [Com (2,191) = 9.367, \( p = .000 \)], was found to be significantly different. In relation to the study area, Humanities’ results are slightly lower than Science and Social Sciences (mean: -0.023). With the aim of exploring the exact nature of the differences observed, we present the percentages related to generic competencies grouped by area of study in the highest five point scale (Table 6).

### Table 6. Incidence of area of study variable in the development of generic competencies

<table>
<thead>
<tr>
<th>Generic Competencies</th>
<th>Arts &amp; Humanities</th>
<th>Science</th>
<th>Social Sciences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis and Synthesis</td>
<td>38.14%</td>
<td>45.13%</td>
<td>41.86%</td>
</tr>
<tr>
<td>Communicative</td>
<td>29.11%</td>
<td>35.02%</td>
<td>37.87%</td>
</tr>
<tr>
<td>Digital</td>
<td>25.71%</td>
<td>32.11%</td>
<td>26.18%</td>
</tr>
<tr>
<td>Foreign Language</td>
<td>39.87%</td>
<td>31.13%</td>
<td>39.02%</td>
</tr>
<tr>
<td>Mechanical</td>
<td>19.86%</td>
<td>43.25%</td>
<td>37.14%</td>
</tr>
<tr>
<td>Information Processing</td>
<td>38.14%</td>
<td>42.07%</td>
<td>43.51%</td>
</tr>
<tr>
<td>Technical</td>
<td>21.12%</td>
<td>46.11%</td>
<td>39.75%</td>
</tr>
<tr>
<td>None</td>
<td>9.57%</td>
<td>13.00%</td>
<td>12.93%</td>
</tr>
</tbody>
</table>

Note. Cronbach \( \alpha = 0.87 \); 5 items; four-point Likert scale (1–5); mean = 5; standard deviation = 0.43; \( N = 419 \).

The highest impact of DMD in all areas of study was on “Analysis and synthesis” (41.69%), “Information processing” (41.24%), and “Foreign language” (36.67%) competencies. The factor-based structure obtained coincides with the theoretical model of the factors that influence innovation with ICT outlined in the framework of this research (Moran, Hawles, & Gayar, 2010; Pachler, Bachmair, & Cook, 2010; Goral, 2011; Eichenlaub et al., 2011; Yoiro, & Feifei, 2012).

These results need to be confirmed according to the DMDs' incidence on developing learning activities and generic competencies from EHEA principles through a structural equation modeling technique. The LISREL 9.1 program was employed for this purpose. The construct validity obtained provides information regarding the different forces that influence in some way how students react to and use DMD in Higher Education. Seven fit indexes which are commonly used in the literature (\( \chi^2/d.f. \), GFI, AGFI, NNFI, CFI, RMSR, RMSEA) were employed to test model fit. The commonly used measures of model fit, based on results from an analysis of the structural model, are summarized in Table 7.

### Table 7. Summary statistics of model fit

<table>
<thead>
<tr>
<th>Fit Index</th>
<th>Recommended value</th>
<th>Observed value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-square/ degrees of freedom</td>
<td>( \leq 3.00 )</td>
<td>2.878</td>
</tr>
<tr>
<td>GFI</td>
<td>( \geq 0.90 )</td>
<td>0.97</td>
</tr>
<tr>
<td>AGFI</td>
<td>( \geq 0.80 )</td>
<td>0.95</td>
</tr>
<tr>
<td>NNFI</td>
<td>( \geq 0.90 )</td>
<td>0.94</td>
</tr>
<tr>
<td>CFI</td>
<td>( \geq 0.90 )</td>
<td>0.97</td>
</tr>
<tr>
<td>RMSR</td>
<td>( \geq 0.10 )</td>
<td>0.07</td>
</tr>
<tr>
<td>RMSEA</td>
<td>( \leq 0.06 )</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Note. GFI = goodness-of-fit index; AGFI = adjusted goodness-of-fit index; NNFI = non-normed fit index; CFI = comparative fit index; RMSR = root mean square residual; RMSEA = root mean square error of approximation.

In practice, Chi-square / degrees of freedom less than 3, GFI, NNFI, CFI greater than 0.9, an AGFI greater than 0.8, RMSR less than 0.1, and RMSEA less than 0.06 or 0.08 are considered indicators of good fit. As seen in Table 7, all goodness-of-fit statistics are in the acceptable ranges. Figure 2 shows the standardized LISREL path coefficients with their respective significance levels. The proposed structural model explained 61% variance in the impact of DMD in Higher Education Spanish students.

Hypothesis 1 and 2 are postulated that DMD could have a positive effect on the development of learning activities and generic competencies. The direct path DMD-Competencies is significant since the regression coefficient (\( \beta \)) is 0.67 with \( t \) value of \( t = 11.41 \), and \( p < 0.05 \). Therefore the hypothesis is supported, which means that DMD significantly has a positive direct effect on different competencies: “Analysis and Synthesis” (C1 .899), “Foreign
language” (C4 .877), “Information processing” (C6 .845). Moreover the direct path DMD-Learning activities is significant since the regression coefficient (β) is 0.71 with t value of t = 11.49 and p < 0.05 with a higher incidence on: “Academic activities” (LA1 .689), “Teacher's explanations” (LA2 .677), and “Didactic materials” (LA3 .847). These results support other studies developed with “Ipads” (Weider, 2011); institutional reports about mobile learning (UNESCO, 2013), and mobile communication in Higher Education (Mercier & Higgins, 2013). The third hypothesis is also accepted because the direct path DMD-PE-EE-SE is significant since the regression coefficient (β) is 0.51 with t value of t = 2.71 and p < 0.05 (PE); (β) is 0.57 with t value of t = 2.82 and p < 0.05 (EE) and (β) is 0.63 with t value of t = 2.83 and p < 0.05 (SE). Spanish university students' performance and effort expectancy and their perception about self efficacy when using DMD are high. The results showed that DMD in higher educational institutions exhibited the strongest direct impact on the development of learning activities and generic competencies what is important to implement new ways of teaching and learning based on ubiquitous and mobile learning with the support of digital mobile devices and apps.

![Figure 2. LISREL test of Research model](image)

**Conclusions**

Without doubt, ubiquitous learning forms a new educational paradigm that stems from new media and technology resources based on the principles of mobility, collaboration, and active participation. This provides alternative learning interactions and access to a great variety of contents and resources. The results from this research in three Spanish universities set one of the first maps of factors that condition the uses and incidence of DMD in EHEA.

The study presented here had a twofold objective. First, it aimed to identify the factors that foster innovation with DMD among students at university, through the development of a scale and the subsequent analysis of its psychometric properties, such as validity and reliability. Secondly, it aimed to compare the different factors that condition DMD adoption according to EHEA principles and generic competencies. In relation to the first objective, the results obtained during the construction and validation of the Scale of Factors that Foster Innovation with DMD (SFFIDMD) among university students was satisfactory. According to the reliability coefficients obtained for both the scale on the whole and each of its dimensions, the instrument was found to be a reliable tool for measuring the factors that influence the promotion of DMD innovation among university students.
Innovation should be considered a value at university, a characteristic that sets it apart from the rest. In this sense, innovation with DMD requires the establishment of communications channels and opportunities for thought, reflection, and collaboration, to foster the effective and efficient engagement of the teaching staff, management board, and students. To this end, it is vital that universities express and reflect this DMD innovation-oriented philosophy in syllabuses. However, an adequate university context alone is not enough to ensure ubiquitous and mobile innovation with ICT. From their hypothetical distance, education authorities also play a key role and have much to offer. This second key element interacts with the first one in a number of basic aspects, such as the provision of resources and infrastructures, teacher training, guidance, and counseling. The final key element is the teachers and professors themselves, who are the key figures and direct executers of any educational innovation in their subjects. Certain beliefs and attitudes among teachers in relation to ICT and its implementation in teaching-learning processes foster innovation in this field. Professional responsibility based on a positive, practical view of ICT as a helpful tool in the classroom fosters innovation, provided that teachers do not go to the other extreme and start viewing each new technological breakthrough (especially those related to information processing) as the ultimate answer to the problems of university-based teaching and learning. It is vital to have a model that aims to guarantee the integration and use of and innovation with DMD by means of a new way of understanding teaching-learning and management processes that go beyond the mere provision of technological resources.

The study suggests that 61% of the variance in the model is explained by students' perception of DMD as useful devices for promoting the acquisition of “competencies” and for developing “learning activities.” Academic organizations adopting ICT innovations will have a much higher probability of success if they provide and require specific training for the proper use of DMD. There are positive correlations between “Performance and effort expectancy, and self efficacy” promoting an increase in the development of “competencies” and “learning activities.” To apply the results of the study to the technology implementation process, the success of any technical innovation appears to hinge on quality planning. Adequate planning should include two components that will improve the success of the initiative. First, there should be an opportunity for all key stakeholders to meet and discuss the initiative. Second, there has to be adequate provision of educational opportunities to provide the skills necessary to use these devices correctly.

We conclude that DMD with high penetration among the age group studied (18-26 years) can enable college students to improve several indicators of generic competencies, especially those of “self-regulated learning,” “higher cognitive,” “communication,” “instrumental in the knowledge society,” and “interpersonal” (Dublin Descriptors, 2004).

Study limitations

The findings of this study should only be applied to this unique environment. The conclusions must be carefully evaluated before any attempt is made to project these findings on another university setting. Student populations in other university environments may very well have very different model analysis and distributions than this study’s participant survey.

Acknowledgments

This work is funded by the Spanish Ministry of Education and it is part of a research, development and innovation project entitled: Ubiquitous Learning with mobile devices: design and development of a map of higher education responsibilities. EDU2010-17420-Sub EDUC.

References


Aptitude-Treatment Interactions during Creativity Training in E-Learning: How Meaning-Making, Self-Regulation, and Knowledge Management Influence Creativity

Yu-chu Yeh* and Chun Fu Lin

Institute of Teacher Education, National Chengchi University, No. 64, Zhihan Road, Section 2, Taipei 116, Taiwan, ROC; Research Center for Mind, Brain & Learning, National Chengchi University, No. 64, Zhihan Road, Section 2, Taipei 116, Taiwan, ROC // Department of Engineering Science and Ocean Engineering, National Taiwan University, No. 1, Roosevelt Road, Section 4, Taipei, 106, Taiwan, ROC // yeyeh@nccu.edu.tw // lcf73211@gmail.com

*Corresponding author

(Submitted December 13, 2013; Revised April 19, 2014; Accepted May 23, 2014)

ABSTRACT

The goal of aptitude-treatment interactions (ATIs) is to find the interactions between treatments and learners’ aptitudes and therefore to achieve optimal learning. This study aimed at understanding whether the aptitudes of meaning-making, self-regulation, and knowledge management (KM) would interact with the treatment of 17-week KM-based training and then influence creativity in e-learning. The participants were 31 university students, and all variables were measured using online systems. ATIs and mediation effects during the training were found. Specifically, while meaning-making indirectly influenced creativity via KM, self-regulation influenced creativity both directly and indirectly via KM; moreover, university students with higher level of KM and self-regulation ability benefited more from the training than their counterparts. This study not only sheds lights on understanding how ATIs influence creativity learning, but also provides a new approach—KM-based training—to improve university students’ creativity in environments of e-learning.

Keywords

Creativity, Aptitude-treatment Interactions, Knowledge Management, Meaning-making, Self-regulation

Introduction

Creativity is the foundation of human civilization (Dietrich & Kanso, 2010). Given the central importance of creativity and the dominance of e-learning in higher education, it is important to identify the underlying mechanisms that contribute to university students’ learning of creativity in e-learning environments and, accordingly, to design effective training programs to enhance their creativity. According to aptitude-treatment interactions (ATIs), individuals differ in their readiness to profit from a particular treatment and individuals may adapt their situations to fit their own characteristics; therefore, finding the interactions between treatments and learners’ aptitudes helps to create an environment in which the treatments match the aptitude of the learner and, further, to achieve optimal learning (Yeh, 2012b). It has also been suggested that the effects of any learning environment on behavioral engagement in learning are mediated by learner characteristics (Sha, Looi, Chen, Seow, & Wong, 2012). Accordingly, designing an e-learning program to improve university students’ creativity is related not only to technological issues but also to learners’ aptitudes.

This study is concerned with the ATIs effects of three aptitudes on the learning of creativity: meaning-making ability, self-regulation ability, and knowledge management (KM) ability. People with high level of meaning-making ability can actively reappraise events or series of events (Davis & Nolen-Hoeksema, 2009). People with great self-regulation ability tend to actively participate in the learning process in terms of behaviors, motivation, and metacognition (Lee, Lim, & Grabowskig, 2009). Finally, people with great knowledge management ability are competent in knowledge acquisition, knowledge sharing, knowledge application, and knowledge creation (Gagné, 2009; Yeh, 2012a). These aptitudes may contribute to the development of creative ideas and the coping of frustration during creative processes. Therefore, this study aimed to investigate whether the aptitudes of meaning-making, self-regulation, and KM would interact with the treatment of creativity training and then influence creativity in e-learning.
Definitions of creativity

A recent consensus of creativity is that creativity refers to the ability of producing responses that are novel and appropriate (Shamay-Tsoory, Adler, Aharon-Peretz, Perry, & Mayseless, 2011). However, whether creativity is a domain-specific or a domain-general ability remains a debate. While some researchers (e.g. Baer & Kaufman, 2005) argued that creativity is a domain-general trait, some researchers claimed that creativity is domain-specific (e.g. Simonton, 2012; Reiter-Palmon, Illies, Cross, Buboltz & Nimps, 2009), and the others (e.g. Silvia, Kaufman, & Pretz, 2009) supported hybrid models in which general factors are required for the development of creativity and domain-specific factors are critical to certain creative activities. This study supports the hybrid model and suggests that creativity is a process in which one generates a culturally “original” and “valuable” response or product within a certain domain. Moreover, during the creation process, both general factors and domain-specific factors are required and a creative outcome is the result of the interactions of personal characteristics and the environment. Accordingly, three personal characteristics that may influence creation process were included in the experimental instruction in this study. However, since the experimental instruction were integrated into a liberal education course, participants were from different disciplines, only the general factors of creativity were investigated in this study.

Over the past six decades, divergent thinking tests have been the most popular evaluation instrument for understanding the general factors of creativity. Such tests measure the ability to generate new ideas, allowing the development of multiple solutions to a given problem. Two central indices of divergent thinking test are fluency and originality. While fluency describes the productivity of ideas, originality refers to the uniqueness of responses (Shamay-Tsoory et al., 2011). These two indices of creativity were measured in this study.

KM and creativity

Most existing theories of KM emphasize the competencies of knowledge acquisition and storage, knowledge application, knowledge sharing, and knowledge creation (Yeh, 2012a). Recently, many information technology industries have integrated blended KM models into their human resources training programs; most models emphasize either knowledge sharing (e.g., Alony, Whymark, & Jones, 2007; Gagné, 2009) or knowledge creation (e.g., Imani, 2007; Yeh, Yeh, & Chen, 2012).

Yeh, Huang, & Yeh (2011) found that KM processes were effectively facilitated in a blended learning environment using the strategies of socialization (e.g., building a learning community, engaging in observational learning, and participating in online group discussions), externalization (e.g., sharing opinions in online discussions and group assignments), combination (e.g., creating stories and design products), and internalization (e.g., giving feedback on performance and providing opportunities for practice). Moreover, many researchers have claimed that knowledge sharing is a key component of KM systems; the influential factors of these KM processes include individual factors, organizational factors, and technological factors (Park, Ribiere, & Schulte, 2004; Riege, 2005). Another key component of KM is knowledge creation, which can be enhanced through shared experiences in social interactions (Nonaka & Toyama, 2003); community building (Swirski, Wood, & Solomonides, 2008), practice, reinforcement, and imitation (Leroy & Ramanantsoa, 1997). Obviously, collaborative knowledge construction as well as the interplay between individual and collective knowledge building are greatly emphasized in KM. Recently, the building of learning community has been regarded as an important mechanism for achieving collaborative knowledge construction. In a well-developed learning community, learners collaboratively communicate during their educational experience to construct knowledge, and such process are often built upon social presence (participants seem like actual people), teaching presence (the design and development of learning experiences), and cognitive presence (the ability of learners to use online communication to construct meaning) (Kucuk & Sahin, 2013).

Studies from e-learning also support the interplay between individual and collective knowledge construction. For example, Moskaliuk, Kimmerle, & Cress (2009) suggested that the cognitive systems of the individuals involved and the social system wiki mutually influence each other, and new knowledge is generated through the processes of discussion, internalization, and externalization. Cress, Held, and Kimmerle (2013) also claimed that tag clouds generated in social tagging systems can capture the collective knowledge of communities, and both collective and individual knowledge have a significant influence on link selection, incidental learning, and information processing.
A sound knowledge base is essential for the development of creativity. Du Plessis (2007) claimed that KM allows collaboration, knowledge sharing, and continual learning, explaining that it plays the following roles in creativity: (1) enabling the sharing of tacit knowledge; (2) making explicit knowledge available for producing creative ideas; (3) enabling the transfer of tacit knowledge via collaborative processes; and (4) conducing knowledge sharing and creation, as well as collaboration, through the creation of a culture. Accordingly, KM is critical to creativity. This study attempted to emphasize collaborative knowledge construction as well as the interaction between individual and collective knowledge building in a training program to facilitate university students’ KM and, thereby, to improve their creativity. Our first hypothesis was proposed as follows:

**H1**: KM would have positive effects on improvements of creativity in KM-based training. Those students with better KM capacities would show more improvement in creativity than their counterparts.

**Meaning-making, KM, and creativity**

Meaning-making refers to the active process through which individuals reappraise an event or a series of events. Such processes usually involve two aspects: benefit-finding and sense-making (Baumeister, 2005; Davis & Nolen-Hoeksema, 2009). Benefit-finding involves the process of transforming adversity into prosperity. A benefit-finding individual tries to find the positive aspects of a negative event. On the other hand, sense-making involves looking for ways to understand the event (Davis & Nolen-Hoeksema, 2009). Klein, Moon, and Hoffman (2006) defined this process as “how people make sense out of their experience in the world” (Davis & Nolen-Hoeksema, 2009).

Regarding the relationship between meaning-making and KM, the data-information-knowledge-wisdom (DIKW) hierarchy provides supporting arguments. The DIKW hierarchy includes four components: data (raw facts), information (relevant contexts), knowledge (meaningful interrelations), and wisdom (perceiving outcomes and determining their values) (Müller & Faltin, 2011). In reflective learning, knowledge results from meaningful interrelations between the information about a reflected situation and information about its environment as well as the feelings and behaviors of the reflecting person (Müller & Faltin, 2011). Moreover, according to the generative learning model (Wittrock, 1992), to comprehend a complex topic, learners must generate meaning for events by constructing relationships between new information and previously acquired information, conceptions, and background knowledge. Therefore, generative learning that emphasizes the actual creation of relationships and meanings is, in essence, knowledge generation (Lee et al., 2009). In other words, meaning-making and KM are closely related during knowledge generation.

As for the relationship between meaning-making and creativity, benefit-finding is a construct that captures lifestyle and behavioral changes in perception (Lenchner, Tennen, & Affleck, 2009). Although benefit-finding is idiosyncratic, most benefits that are reported following adversity can be categorized within the areas of relating to others, new life possibilities, personal strength, appreciation of life, and spiritual or religious change (Lenchner et al., 2009). Sense-making, based on its characteristics of enactment and plausibility, can be viewed as the reciprocal interaction, including seeking cues, assigning meaning, and moving towards plausible action. The extracted cues from one's environment (e.g., availability of resources for creativity) may act as triggers or may signify that certain meaning is required (Madjar, Nora, Greenberg, & Chen, 2011). Moreover, sense-making may represent curiosity; it has been referred to as the trigger for “scientific imagination” (Klein et al., 2006). Thus, the positive dispositions of benefit-finding and sense-making may contribute to the development of creative ideas and coping with frustration during creative processes.

Integrating the relationship between meaning-making, KM, and creativity, we proposed the following two hypotheses:

**H2**: Meaning-making would have positive effects on the improvement of creativity. Those students with better meaning-making abilities would show more improvement in creativity than their counterparts in creativity training.

**H3**: Meaning-making would have positive effects on KM. Those students with better meaning-making capacities would show better performance in KM than their counterparts in creativity training.
Self-regulation, KM, and creativity

Self-regulated learners are active participants in their own learning processes in terms of behaviors, motivation, and metacognition (Lee et al., 2009). The various self-regulation learning (SRL) theories share three basic assumptions, stating that self-regulated learners are able to (1) personally improve their abilities to learn through the selective use of metacognitive and motivational strategies; (2) proactively select, organize, and even create advantageous learning environments; and (3) play significant roles in choosing the forms and amounts of instruction they need (Sha et al., 2011). Recently, Fruhmann, Nussbaumer, and Albert (2010) proposed the Responsive Open Learning Environments (ROLE) model, in which self-regulated learning is defined by four learner-centered phases: (1) learner profile information is defined or revised; (2) the learner finds and selects learning resources; (3) the learner works on selected learning resources; and (4) the learner reflects on and reacts to strategies, achievements, and usefulness. These phases are summarized in the “plan-learn-reflect-plan” loop.

As for the relationship between self-regulation and KM, Lee et al. (2009) proposed that SRL guides generative activities that enhance learners’ knowledge generation by creating relationships between new information and prior knowledge. In addition, KM and technology-enhanced learning are confronting many new challenges due to the rapid pace of technological progress (Müller & Faltin, 2011). Müller and Faltin (2011) argued that in work-intensive environments, a formal method of knowledge acquisition and learning is often insufficient. Therefore, a new generation of tools supporting self-regulated learning is needed.

Regarding the relationship between self-regulation and creativity, research findings have suggested that self-perceived creativity is positively related to daily planning, confidence in long-range planning, overall time management, perceived time management, and tenacity and negatively related to preference for disorganization. Zampetakisa, Bourantab, and Moustakis (2010) claimed that time management skills, which characterize self-regulated learners, are related to creativity. In related studies, Hon (2011) found that self-concordance mediated social-contextual variables and creativity. Self-concordant goals are intrinsically motivating because they are derived from self-choice. People with high levels of self-concordance are, therefore, competent in identifying and pursuing goals (Hon, 2011), which is critical for self-regulated learners. Along the same lines, King and Gurland (2007) found that autonomous orientation is related to the detail and complexity aspect of creativity. Creative individuals tend to overcome barriers by deploying time management skills that maximize effectiveness as a function of time; such a tendency can also be seen in self-regulated learners, who develop plans and strategies and monitor their behaviors to attain their anticipated goals (Zampetakisa et al., 2010). Accordingly, self-regulation should contribute to the learning of creativity.

Integrating the relationship between self-regulation, KM, and creativity, we proposed the following two hypotheses:

**H4:** Self-regulation would have positive effects on the improvement of creativity. Those students with better self-regulation abilities would show more improvement in creativity than their counterparts in creativity training.

**H5:** Self-regulation would have positive effects on KM. Those students with better self-regulation capacities would show better performance in KM than their counterparts in creativity training.

**Method**

**Participants**

The participants were 31 undergraduates (9 males and 22 females) with a mean age of 19.93 years (SD = 1.44 years). All of the participants were enrolled in a liberal education course “Creativity,” which emphasized KM and e-learning.

**Instruments**

The instruments employed in this study included an e-learning website (http://moodle.nccu.edu.tw/) and an online experimental system developed by PHP and JavaScript. The experimental system included the Inventory of
Knowledge Management in E-learning (IKME), the Inventory of Meaning-making in E-learning (IMME), and the Inventory of Self-regulation in E-learning (ISRE).

The Digital Imagery Test, a divergent thinking test, was employed to test creativity in this study. The Digital Imagery Test, which includes 12 pictures, was developed via the PHP programming language based on a picture book (Shaw, 1993) using Adobe Flash Professional CS4 (Yeh, 2011) (see Figure 1). The pictures were displayed in a fixed order on the website for 2 minutes each. The participants were encouraged to imagine what they had seen and then type in as many answers as possible. The Digital Imagery Test scores included two commonly measured indices of creativity: fluency and originality (Mayer, 1999). When an answer was appropriate, the fluency is scored as “1” point; the total fluency score was the sum of appropriate answers of all picture. On the other hand, given that X is the percentage of an answer, the scoring of originality was determined as follows: when $X \geq 16\%$, originality was scored “0” points; when $5\% \leq X < 16\%$, originality was scored “1” point; when $2\% \leq X < 5\%$, originality was scored “2” points; and when $X < 2\%$, originality was scored “3” points. The test-retest reliability coefficients based on a 3 month interval were .779 and .785 ($p < .001$) for fluency and originality. The total originality score was the sum of appropriate answers of all pictures. The Digital Imagery Test has good criterion-related validity, the ability of fluency and originality are positively related self-evaluation of creativity, $r_{s}(33) = .413$, and .433, $p < .05$, respectively.

![Figure 1. A sample screenshot from the Digital Imagery Test](image)

All three e-learning inventories were validated by reliability analyses ($N = 1017$), exploratory factor analyses ($N = 1017$), and confirmatory factor analyses (CFA) ($N = 1647$) (Yeh, Yeh, & Lin, 2013). Moreover, all e-learning inventories used 6-point Likert scales, with response options ranging from “totally disagree” to “totally agree.” The tests were administered online with no time limit imposed.

The IKME, with a total of 22 items, included four factors: knowledge acquisition (7 items), knowledge application (6 items), knowledge sharing (5 items), and knowledge creation (4 items). Example items include “I participate in or organize e-learning communities (e.g., BBS, Facebook, etc.) to increase interactions with others” and “I try to integrate the knowledge I have learned to produce creative ideas during e-learning.” The Cronbach’s alphas for the four factors and the IKME were 0.887, 0.897, 0.827, 0.910, and 0.942, respectively.

The IMME, with a total of 25 items, included two factors: benefit-finding (14 items) and sense-making (11 items). Sample items include “although conducting discussions in e-learning platforms is time consuming, it stimulates my multi-perspective views and creative ideas” and “I can’t effectively interact with others in an e-learning interface because I am not used to the interface”. The Cronbach’s alphas for the two factors and the IKME were 0.922, 0.869, and 0.884, respectively.

The ISRE, with a total of 16 items, included three factors: information retrieval and integration (6 items), strategy adaptation and progress monitoring (7 items), and time and efficiency management (3 items). Example items include “when conducting e-learning, I can adjust my methods of searching resources to find useful information” and “when conducting e-learning, I plan my learning time”. The Cronbach’s alphas for the three factors and the ISRE were 0.882, 0.872, 0.793, and 0.924, respectively.
**Experimental design and procedures**

This study employed a before-and-after design. A 17-week creativity training program, which emphasized the integration of KM processes and e-learning, was designed to investigate the relationships between meaning-making, self-regulation, KM, and improvements in creativity in e-learning environments. The pretest was administered during the second week, while the posttest was administered at week 17. The pretests included the Digital Imagery Test, the IKME, the ISRE, and the IMME, whereas the posttests included the Digital Imagery Test and the IKME.

Systematic lectures were given in combination with in-class and online discussion activities throughout the training program. The goal of the program was to enhance participants’ creativity (including fluency and originality) by facilitating KM processes. Based on past findings (Cress, Held, & Kimmerle, 2013; Kucuk & Sahin, 2013; Nonaka & Toyama, 2003; Park et al., 2004; Riege, 2005; Swirski et al., 2008; Yeh et al., 2011, Yeh 2012a), several strategies were incorporated into the training program. To enhance knowledge acquisition, online information was requested (e.g., searching for creative products, creative games, and creativity evaluation methods). To facilitate knowledge sharing, the following activities were emphasized: (1) building learning communities through self-determinate grouping and group discussions; (2) practicing observational learning via in-class presentations of group assignments; and (3) online sharing and evaluations of other groups’ assignments. To facilitate knowledge application, abundant practice in creative strategies, discussions, and interactions were provided. Finally, knowledge internalization, group discussions, and the design of creative products were used to achieve knowledge creation.

Specifically, the following seven assignments were given and scored in this study: participation in in-class discussion and instructional activities (including lectures, discussions, presentations of assignments), online and in-class sharing of creative products, online and in-class sharing of creative games, and four assignments for producing creative products (creative self-introduction, mind maps, stories of positive thinking, and digital creative story-telling). Throughout the training program, e-learning emphasizing the integration of online learning and in-class activities was employed. Specifically, all assignments were completed and shared online for discussion after in-class lectures and practices. Then, in-class discussions followed the online discussion that had lasted for one week. The online discussion allows learners to go beyond the time and space constraints and provides support for community building (Gao, 2011) which is important to knowledge management (Kucuk & Sahin, 2013; Yeh, 2012a). Comparatively, the face-to-face discussion is more prompt in responses and more multidirectional in interaction (Wang & Woo, 2007), which contributes to knowledge integration, knowledge creation, and knowledge internalization (Yeh et al., 2012).

![Figure 2. The hypothesized model and instructional design of this study (H1 to H5 represent hypothesis 1 to hypothesis 5)](image-url)
Based on the aforementioned literature review, meaning-making and self-regulation may directly influence the learning of creativity as well as indirectly influence the learning of creativity via KM in an e-learning environment. Thus, this study attempted to improve university students’ creativity through facilitating KM in an e-learning program. The hypothesized model and the instructional design are illustrated in Figure 2.

Data analysis

A two (within-group variables: pretest vs. posttest) by two (between-groups variable: high vs. low) mixed design of repeated measures Analysis of Variance (repeated measures ANOVA) was used to examine whether the three positive personal traits (KM, meaning-making, and self-regulation) influenced improvements in creativity. Moreover, univariate ANOVAs were employed to investigate the effects of meaning-making and self-regulation on KM. The between-groups variables of the three positive personal traits were divided into high and low groups based on the median score. Because more than one participant obtained the median score, the group sizes differed.

Results

Preliminary analyses

Preliminary analyses of the correlations between important interventions and creativity performance in this study found that (1) the frequency of online discussion and the performance of mind mapping were related to the performance of creative digital story-telling—the most important group assignment in this study, \( r = 0.395, p < 0.05 \) and \( r = 0.612, p < 0.001 \); (2) the frequency of online discussion was related to the posttest scores of fluency and originality measured by the Digital Imagery Test, \( rs = 0.445 \) and 0.446, \( ps < 0.01 \); (3) self-perceived knowledge-creation measured by the IKME was related to the performance of creative digital story-telling, \( r = 0.360, p < 0.05 \); (4) participation in in-class discussion and instructional activities was related to the performance of creative digital story-telling and mind mapping, \( rs = 0.460 \) and 0.486, \( ps < 0.01 \); (5) The composite score of the seven major assignments was related to the posttest score of fluency and originality, \( rs = 0.526 \) and 0.521, \( ps < 0.01 \).

These findings revealed that the creativity measured by the divergent thinking test of Digital Imagery Test was closely related to the participants’ product-oriented creativity, suggesting that the score of Digital Imagery Test is predictive to actual creativity performances; moreover, the close relationships between interventions, knowledge creation, and creativity performances suggest that the interventions are effective in improving the participants’ creativity.

The effects of KM on improvements in creativity

To investigate whether the training program would enhance the influence of KM on improving creativity, we separately analyzed the influences of pretest KM scores and of posttest KM scores on creativity improvement. Figure 3 (a) and Figure 3 (b) display the mean creativity scores in the different KM groups. Using the pretest KM group as the independent variable, the 2 (pretest KM group: high vs. low) × 2 (test: pretest vs. posttest) repeated measures ANOVA revealed that the interaction effect was not significant for fluency or originality. However, the main effect of the test (pretest vs. posttest) was significant for both fluency and originality, \( F(1, 29) = 28.464, p < 0.001, \eta^2_p = 0.479 \) and \( F(1, 29) = 24.670, p < 0.001, \eta^2_p = 0.443 \), respectively. The main effect of the group was also significant for fluency and originality, \( F(1, 29) = 5.698, p = 0.023, \eta^2_p = 0.155 \) and \( F(1, 29) = 3.996, p = 0.054, \eta^2_p = 0.114 \), respectively.

Using the posttest KM group as the independent variable, the 2 (posttest KM group: high vs. low) × 2 (test: posttest vs. pretest) repeated measures ANOVA analysis revealed that the interaction effect was not significant for fluency or originality. However, the main effect of the test was significant for both fluency and originality, \( F(1, 29) = 28.675, p < 0.001, \eta^2_p = 0.497 \) and \( F(1, 29) = 23.972, p < 0.001, \eta^2_p = 0.453 \), respectively. The main effect of the group was also significant for fluency and originality, \( F(1, 29) = 9.185, p = 0.005, \eta^2_p = 0.241 \) and \( F(1, 29) = 10.590, p = 0.003, \eta^2_p = 0.267 \), respectively.
In both analyses, comparisons of the means revealed that the participants had higher fluency and originality scores on the posttest than on the pretest and that those participants with higher levels of KM ability showed greater improvements in fluency and originality than those with lower levels. Moreover, the relationship between posttest KM and improvements in creativity was much stronger than that between pretest KM and improvements in creativity.

Figure 3. Mean creativity scores in different KM groups

The effects of meaning-making on improvements in creativity

Figure 4 (a) displays the mean scores for creativity in the different meaning-making groups. The 2 (group: high vs. low) × 2 (test: pretest vs. posttest) repeated measures ANOVA analysis revealed that the interaction effect was not significant for fluency or originality. Moreover, the main effect of the group was not significant for fluency or originality. However, the main effect of the test was significant for both fluency and originality, \( F(1, 29) = 27.588, p < 0.001, \eta_p^2 = 0.488 \) and \( F(1, 29) = 22.950, p < 0.001, \eta_p^2 = 0.442 \), respectively. Comparisons of the means revealed that the participants obtained higher scores for fluency and originality on the posttest than on the pretest, but those participants with higher levels of meaning-making ability did not show greater improvements in fluency or originality after the training compared with those participants who had lower levels of meaning-making ability.

Figure 4. Mean creativity scores in the different meaning-making and self-regulation groups
The effects of self-regulation on improvements in creativity

Figure 4 (b) displays the mean creativity scores in the different self-regulation groups. The 2 (group: high vs. low) × 2 (test: pretest vs. posttest) repeated measures ANOVA analysis revealed that the interaction effect was not significant for fluency or originality. However, the main effect of the test was significant for fluency and originality, \( F(1, 29) = 27.581, p < 0.001, \eta^2_p = 0.471 \) and \( F(1, 29) = 24.067, p < 0.001, \eta^2_p = 0.437 \), respectively. The main effect of the group was also significant for fluency, \( F(1, 29) = 6.184, p = 0.018, \eta^2_p = 0.166 \) and originality, \( F(1, 29) = 6.327, p = 0.005, \eta^2_p = 0.170 \). Comparisons of the means revealed that the participants had higher scores for fluency and originality on the posttest than on the pretest, and those with higher levels of self-regulation showed greater improvements in fluency and originality than those with lower levels of self-regulation.

The effects of meaning-making and self-regulation on KM

With the total meaning-making score as the independent variable and the total KM score as the dependent variable, the Univariate ANOVA analysis revealed that meaning-making had a significant effect on KM, \( F(1, 29) = 7.167, p = 0.012, \eta^2_p = 0.198 \), suggesting that those participants with greater meaning-making ability were more competent in KM than their counterparts.

Using the total self-regulation score as the independent variable and the total KM score as the dependent variable, the Univariate ANOVA analysis revealed that self-regulation had a significant effect on KM, \( F(1, 29) = 22.445, p < 0.001, \eta^2_p = 0.420 \), suggesting that those participants with greater self-regulation were more competent in KM than their counterparts.

Discussion

Effectiveness of the training program

This study proposed five hypotheses when investigating whether the KM-based training program designed in this study would strengthen the relationship between KM and creativity and whether KM would mediate meaning-making, self-regulation, and the learning of creativity in such a program. Except for the hypothesis concerning the direct influence of meaning-making on the learning of creativity, all of the proposed hypotheses were supported.

The findings in this study revealed that, after the training, all participants significantly improved in creativity, as evidenced by the significant improvement in scores of fluency and originality after the training as well as (see the repeated measures ANOVA) the close relationships between the major interventions, knowledge creation and creativity performances (see the preliminary analyses). Moreover, the KM-based training successfully strengthened the influence of KM on creativity and, therefore, enhanced the improvements in creativity. Since the main purpose of this study was to investigate the relationship between KM, mediate meaning-making, and self-regulation during the learning of creativity in a 17-week training program, the control group was not employed. Nevertheless, the findings based on a deliberately designed program can provide valuable information for the teaching of creativity.

The training program is, in essence, conducted in an e-learning environment that emphasizes four KM processes: knowledge acquisition, knowledge sharing, knowledge application, and knowledge creation. Therefore, the findings suggest that university students’ creativity can be enhanced through the facilitation of KM processes in a 17-week e-learning training program. The training program in this study emphasizes the learning community, group discussions, observational learning, online sharing and evaluation, abundant practice and interactions, and group assignments. The effectiveness of this training program suggests that these mechanisms contribute to the facilitation of KM processes and creativity. The findings in this study also support the arguments that creativity is the process of knowledge building (e.g., Craft, 2005; Paavola, Lipponen, & Hakkarainen, 2004) and that KM plays a critical role in the development of creativity (Du Plessis, 2007). Moreover, the findings suggest that KM processes can be effectively facilitated in e-learning environments through the KM strategies of socialization, externalization, combination, and internalization (Cress et al., 2013; Kimmerle, Cress, & Held, 2010; Yeh et al., 2011)
Aptitude-treatment interactions during the training

This study found aptitude-treatment interactions during the training. Specifically, the participants with higher levels of KM ability showed greater improvements in fluency and originality following the training than those with lower levels of KM. Similarly, the participants with greater self-regulation showed greater improvements in fluency and originality after the training than those with less self-regulation.

According to the DIKW model, knowledge is defined as information connected through relationships (Müller & Faltin, 2011). Participants with better KM are, therefore, more capable of connecting information with creativity during training. As a result, they benefit more from training than their counterparts do. Existent theories and previous empirical findings on self-regulation can also be used to describe aptitude-treatment interactions and their effects on training. For example, past studies have suggested that self-regulated learners are more capable of making plans, managing time, identifying and pursuing goals, and conducting autonomous learning (Hon, 2011; King & Gurland, 2007; Zampetakisa et al., 2010). Accordingly, people with greater self-regulation are more competent in managing their learning processes during the learning of creativity, therefore benefitting more than their counterparts.

Mediating effects during the training

Notably, this study also found mediating effects during the training; specifically, meaning-making and self-regulation influenced the improvement of creativity through KM. In addition, self-regulation had stronger indirect effects than meaning-making did ($\eta_p^2 = 0.012$ vs. 0.420). These findings support the aforementioned DIKW theory, in that knowledge in reflective learning is derived from meaningful interrelations between information about a reflected situation and its environment (Müller & Faltin, 2011). These results also support the argument that the creation of relationships and meaning is, in essence, knowledge generation (Lee et al., 2009). The findings in this study are in line with the claims that different social value orientations lead to different perceptions of the costs and benefits of knowledge sharing decisions, thereby influencing the inclination to share knowledge (Cyr & Choo, 2010). Moreover, the decision to engage in creative work entails, in addition to personal and contextual factors, sophisticated cognitive processes and sense-making activities (Madjar et al., 2011). Accordingly, university students with good meaning-making abilities are able to employ KM processes effectively, further enhancing their creativity.

As for the mediating effects of KM on self-regulation and creativity, the findings in this study support the argument that self-regulated learners are able to generate activities that enhance knowledge generation by creating relationships between new information and prior knowledge (Lee et al., 2009); in addition, the results are consistent with suggestions that self-regulated learners are characterized by self-concordance, autonomous orientations, and goal setting (Hon, 2011; King & Gurland, 2007; Zampetakisa et al., 2010). Accordingly, university students who are competent self-regulators should be able to flexibly employ KM strategies to achieve their goals and, further, to improve their creativity.

Conclusions and suggestions

In this epoch of information technology and knowledge economics, KM and creativity have become required competences for university students. This study designed a KM-based training program to facilitate university students’ KM processes and, further, to enhance their learning of creativity. Most importantly, this study mainly aimed at understanding whether university students with different aptitudes would benefit differently from the training program. The findings in this study suggest that KM is a mediator between meaning-making, self-regulation and the learning of creativity in an e-learning environment, and ATIs exist during the training.

Due to the long period of experimental instruction and creativity training that emphasizes discussions and interactions can be better achieved via a small sample, only a small sample was included. However, since the training program in this study is deliberately designed based on a sound base of theories and empirical findings and the training session lasts for 17 weeks, the findings have important implications to the curriculum and instructional design aimed at improving creativity in an e-learning environment. Accordingly, this study not only sheds lights on understanding how ATIs influence creativity learning, but also provides a new approach—KM-based training—to improve university students’ creativity in environments of e-learning.
Due to the great variety of participants’ disciplines, the designed instructional activities in this study were based on a domain-general perspective; further studies can replicate the KM-based framework of instructional design employed in this study to facilitate domain-specific creativity. Moreover, this study found that self-regulation is predictive to the learning of creativity. Nussbaumer, Steiner, and Albert (2008) proposed that the following six self-regulatory processes are important for web-based learning: (1) goal setting supported by communication tools; (2) the use of task strategies supported by content delivery tools (e.g., concept mapping software; (3) self-monitoring supported by the use of discussion forums; (4) self-evaluation supported by the use of rubrics, evaluation criteria, and peer feedback; (5) time planning and management supported by communication tools meant for time budgeting; and (6) help-seeking supported by hypermedia tools. These mechanisms can be taken into consideration while designing a training program for the learning of creativity.

Acknowledgments

This study was supported by the National Science Council of the Republic of China in Taiwan (Contract No. NSC 98-2511-S-004-001-MY2 and NSC 100-2511-S-004-002-MY3).

References


130


Development and Use of an Adaptive Learning Environment to Research Online Study Behaviour

Anna Helga Jonsdottir*, Audbjorg Jakobsdottir and Gunnar Stefansson
Department of Natural Sciences, University of Iceland, Dunhaga 5, 107 Reykjavik, Iceland // ahj@hi.is // audbjorg@yahoo.com // gunnar@hi.is

*Corresponding author

ABSTRACT

This paper describes a system for research on the behaviour of students taking online drills. The system is accessible and free to use for anyone with web access. Based on open source software, the teaching material is licensed under a Creative Commons License. The system has been used for computer-assisted education in statistics, mathematics and fishery science. It offers a unique way to structure and link material, including interactive drills with a purpose of increasing learning rather than mere evaluation. It is of interest to investigate what affects how students learn in such an environment, for example how the system entices students to continue to work. One research question is therefore: When do learners decide to stop requesting drill items? A case study has been conducted including 316 students in an undergraduate course in calculus. Analysis of the data showed that the probability of stopping increases with higher grades but decreases with increased difficulty and the number of questions answered. Also, the probability of stopping decreases if the last question was answered correctly in comparison to when the last question was answered incorrectly.

Keywords
Web-based education, Statistics education, IRT, AIWBES, The tutor-web

Introduction

With the increasing number of web-based educational systems and learning environments several types of systems have emerged. These include the learning management system (LMS), learning content management system (LCMS), virtual learning environment (VLE), course management system (CMS) and Adaptive and intelligent web-based educational systems (AIWBES). The terms VLE and CMS are often used interchangeably, CMS being more common in the United States and VLE in Europe.

The LMS is designed for planning, delivering and managing learning events, usually adding little value to the learning process nor supporting internal content processes (Ismail, 2001). A VLE provides similar service, adding interaction with users and access to a wider range of resources (Piccoli, Ahmad & Ives, 2001). The primary role of a LCMS is to provide a collaborative authoring environment for creating and maintaining learning content (Ismail, 2001). Classes taught on these platforms are accessible through a web-browser but are usually private, i.e., only individuals who are registered for a class have access to the password-protected website.

A number of content providers can be found on the web. Even though they are not educational systems per se, linking them to learning systems would make the content available to a larger audience and save work on creating material within the learning systems. Examples of existing content providers are Khan Academy and Connexions. A number of academic institutions have also made educational material available, including MIT OpenCourseWare and Stanford Engineering Everywhere.

Many systems are merely a network of static hypertext pages (Brusilovsky, 1999) but adaptive and intelligent web-based educational systems (AIWBES) use a model of each student to adapt to the needs of that student (Brusilovsky & Peylo, 2003). These systems tend to be subject-specific because of their structural complexity and therefore do not provide a broad range of content. The first AIWBES systems were developed in the 1990s. These include ELM-ART (Brusilovsky, Schwartz & Weber, 1996; Weber & Brusilovsky, 2001) and the AHA! system (De Bra & Calvi, 1998). Since then, many interesting systems have been developed, many of which focus on a specific subject, often within computer science. Examples of AIWBES systems used in computer science education are SQL-Tutor (Mitrovic, 2003), ALEA (Bieliková, 2006), QuizGuide (Brusilovsky & Sosnovsky, 2005; Brusilovsky, Sosnovsky &
Shcherbinina, 2004) and Flip (Barla et al., 2010) which includes an interesting way of allocating quiz questions to students (discussed further in the following section). AIWBES systems can be found in other fields such language teaching (Chen, Lee & Chen, 2005; Heift & Nicholson, 2001) and teaching modelling of dynamic systems (Zhang et al., 2014) to name some. Systems including competitive web-based drill games are also available, with an overview presented in González-Tablas, Fuentes, Hernández-Ardieta and Ramos (2013).

The goal of the project described here is to build an AIWBES including the functionalities of an LCMS. The system should be open to everyone having access to the web and provide broad educational content including interactive drills with the primary purpose of enhancing learning. Intelligent methods will be used for item allocation in drills and for directing students toward appropriate material. As discussed in Romero and Ventura (2007), great possibilities lie in the use of educational datamining. The behaviour of the students in the system are logged so the system provides a testbed for research on web-assisted education such as drill item selection methods.

It has been described earlier how students tend to strive for higher grades in similar systems (Stefansson, 2004). The present paper considers these drivers more explicitly, namely how the student behaviour, including stopping times, reflects their achievements and likely immediate performance, as predicated by system design.

**Item allocation in educational systems**

Numerous educational systems with different functionality are available today as discussed in the previous section. The majority permits the creation of quiz questions and administration of quizzes for evaluation or to enhance learning. In most systems these quizzes are static, where the instructor has chosen a fixed set of items. In some cases items are selected randomly from an available question pool so that students are not all presented with the same set of questions. In this section, methods for allocating quiz questions or drill items to learners are discussed.

Although there are a number of educational web-based systems that use intelligent and/or adaptive methods for estimating learner’s knowledge in order to provide personalized content or navigation (Barla et al., 2010) only a few systems use adaptive and/or intelligent methods for item allocation (adaptive item sequencing). Even though adaptive item sequencing is not common in educational systems, it has been used in Computerized Adaptive Testing (CAT) (Wainer, 2000) for decades. In CAT the test is tailored to the examinee’s ability level by means of Item Response Theory (IRT).

IRT (Lord, 1980) is the framework used in psychometrics for the design, analysis, and grading of computerized tests to measure abilities. It has been used extensively in CAT for estimating students abilities. Within the IRT framework, several models have been proposed for expressing the probability of observing a particular response to an item as a function of some characteristic of the item and the ability of the student, the Rasch model being a common one (Wright, 1977). Another, slightly more complicated model, is the three parameter logistic model, or the 3PL model, including a difficulty parameter \( \beta \), a discrimination parameter \( \alpha \) and a guessing parameter \( c \). The Point Fisher Information (PFI) is then used to select the most informative item in the pool, that is the item that minimizes the variance in the ability estimate. Using IRT, a test developer is able to have items that can discriminate students along a continuum of the hypothesized latent construct. However, IRT requires a large sample size for item calibration (i.e., getting estimates for the parameters of the model) and thus it is typically not done in the classroom. As an example of a system using this technique is the SIETTE system (Conejo et al., 2004), a web-based testing system (i.e., not used for learning purposes).

Research on the application of IRT in learning environments is largely absent (Wauters, Desmet & Van Den Noortage, 2010). Review of available literature found only one system using IRT for adaptive item sequencing with the main focus on enhancing learning, the web-based programming learning system Flip (Barla et al., 2010) developed within the PeWePro1 project (Bieliková and Návrat, 2009). The system uses, among other methods, IRT to select questions with adequate difficulty using the 3PL model, but the parameters \( (\alpha \) and \( \beta \)) are set manually for each question. Experiments using Flip showed remarkable improvements in test results (Barla et al., 2010).
The tutor-web system has been designed by considering four major requirements:

- to be open source and domain/subject independent
- to provide a wide range of open content through a web browser
- to use intelligent methods for item allocation (and grading), amenable to research
- to function as a LCMS

The system has been in development for several years. A pilot version, written only in HTML and Perl, is described in Stefansson (2004). The new implementation described below incorporates fresh approaches to individualized education. It is written in Plone (Nagle, 2010) which is an open source content management system with a usability focus, written in Python on top of the ZODB object database. Plone is flexible, customisable and extended with packages from a worldwide community. It is popular with educational content providers, powering Connexions (http://cnx.org/), MIT’s OpenCourseWare (http://ocw.mit.edu/) as well as many OpenCourseWare projects, based on the eduCommons (http://educommons.com/) system.

The educational material available within the tutor-web covers wide areas within mathematics and statistics, although only one course is analysed in this study. Examples of use in other fields include fishery science, earth science and computer science. The system could equally well be used in fields as diverse as linguistics, religion, psychology and English. This contrasts several systems which address very special skills, some named above.

The system offers a unique way to structure and link together teaching material and organize it into manageable units both for instructors and students. A well-defined structure enables instructors to construct, share and re-use material and provides a single repository of teaching material for students minimizing time otherwise wasted on imperfect searching and browsing and eradicating any format/incompatibility issues. Additionally, interactive drills have a primary purpose of increasing learning, placing evaluation a distant second. Though the tutor-web is not originally designed as a remote-learning system it can be used as such if desired. The system also provides a testbed for research on web-assisted education such as drill item selection methods.

The tutor-web system is based solely on generic formats and open source software to provide unrestricted usage of material by institutions of limited resources without overheads in terms of software or licenses. The teaching material is licensed under the Creative Commons Attribution-ShareAlike License (http://creativecommons.org/) and is accessible to anybody having a web-browser and web access. Instructors anywhere can obtain free access to the system for exchanging and using teaching material while students have free access to its educational content including self-evaluation in form of drills.

Content structure

The teaching material is organized into a tree (fig. 1) with departments, courses, tutorials, lectures and slides. The different departments can be accessed from the tutor-web homepage (http://tutor-web.net).

The basic component of the tutor-web is the slide. Students can browse through the slides and view related material relevant to a given slide. Instructors can use the slides as source material for presentations in the classroom. Although electronic slides are useful for presentation purposes, often they are not sufficient material when a course is taught. Typically more detailed, examples or even extensive handouts may be needed. Such additional material can be linked to any slide on the tutor-web. As in a classroom, a typical lecture, including several slides, should be constructed around a specific subject to aid and focus the students understanding on that particular subject.

A tutorial typically contains several lectures and should be based on a distinctive topic. A tutorial can belong to more than one course and should be built up around a single theme. For example, a tutorial on simple linear regression could both be a part of a general course on regression and an introductory statistics course. Having the tutorials allows the student to complete a portion of the material and perform self-evaluation based on smaller blocks than a complete course.
The system uniquely uses the modularity and traceability of content so the instructor can easily demonstrate how examples and images are derived: An image based on data can be drawn in the system using the statistical environment R (R Development Core Team, 2011). Normally such an image is presented statically on a screen for a class. Here, however, the R plotting commands and the data are stored as objects in the system, automatically producing PDF or HTML slides. The student can view the underlying data and R code, making the system an ideal tool to teach not only model output but also modelling.

Figure 1. The structure of the tutor-web

One goal of the tutor-web project is to make available a repository of educational material for a BSc degree in mathematics and an MSc degree in applied statistics. Some courses are ready for general use with slides, handouts and questions while others are only placeholders waiting to be filled up with material. In addition to university courses, complete high school mathematics tutorials in Icelandic and English are available with over 2000 drill items.

Drills and item selection

Drills (items) are grouped together so they correspond to material within a lecture. These will be termed “drills” rather than “quizzes” to indicate the emphasis on increasing learning. The drills are multiple choice and the author can choose the number of answers and provide detailed explanation of the correct answer shown to the learners after answering a drill. A drill in the tutor-web system differs from the typical classroom testing methods where a student normally answers a given number of questions during a specific time period. In the tutor-web a student can dynamically start or re-enter a drill, one question at a time and may attempt the drill at his/her leisure, although an instructor might decide on a time limit for recording results.

The intuitive style of the drill in the tutor-web encourages students to improve their results and learn from their mistakes. The learners are provided with knowledge of correct response feedback after answering a drill along with elaborative feedback if provided by the author of the drill. Studies have shown that frequent feedback given to students yields substantial learning gains (Black and Wiliam, 1998). In live applications the students have been encouraged to request answers repeatedly until a decent grade is obtained. Students who do not know the material
can test their knowledge and, upon finding it wanting, go back to the online text or textbooks to come back to the drill at a later date.

In the original version of the tutor-web system, drill items within the same lecture were selected randomly with uniform probability (Stefansson, 2004). In the current version of the system a probability mass function that depends on the grade of the student is used to select items of appropriate difficulty. Here, the difficulty of a question is simply calculated as the ratio of incorrect responses to the total number of responses. The questions are then ranked according to their difficulty, from the easiest question to the most difficult one.

A student with a low grade should have higher probability of getting the lowest ranked questions while a student with a high grade should have higher probabilities of getting the most difficult questions. The mass of the probability function should therefore move towards the difficult questions as the grade goes up. The probability mass functions are shown in figure 2 for a lecture with 100 drill items, based on the implemented discrete forms of the beta distribution. Here, beginning student (with a score 0) receives easy items with high probability. As the grade increases the mode of the probability mass functions shifts to the right until the student reaches a top score resulting in high probability of getting the most difficult items.

Figure 2. Probability mass functions for item allocation in a lecture with 100 questions

Grading

Although the main goal of the quizzes in the tutor-web is learning there is a need to evaluate the student’s performance. The drills are of arbitrary length since the students are permitted to continue to answer questions until they (or the instructor) are satisfied. Because of this, issues regarding how to compute the grade arise.

Currently a student gets one point for answering a question correctly and -1/2 for an incorrect answer. Since the purpose is to enhance learning and thus improve the grade, only the last eight answers are used when the grade is calculated for each lecture. Old sins are therefore forgotten. The student can track the grade and thus monitor personal progress with a single click.

Users and access

Four types of users are defined in the tutor-web: Regular users, students, teachers and managers. There is open access for regular users (anybody having access to the web) to browsing and downloading of material. However, in order to
take drills the user needs to log in to the system and become a tutor-web student. When a user initially signs up, the requirement is to provide a full name, a valid email address, choose a unique user name and agree to data being used for research purposes.

Teachers are editors of tutorials and have the ability to edit and insert material as well as quizzes. They also have access to drill results. Managers have the same privileges as instructors with the additional authority to add and edit Departments and Courses and give teacher rights.

**Viewing educational material**

There are three different ways for a tutor-web user to view the teaching material (fig. 3):

- through a web browser
- as a collection of lecture slides
- as a tutorial printout

The first approach is the simplest one for a student wishing to access the educational material. The student simply needs to open the home-page (http://tutor-web.net) and select a department to see courses containing several tutorials and lectures. The student can then enter a lecture to browse through, slide by slide. As an example, one can enter the Math department, click on Computing and calculus for applied statistics (course), Some notes on statistics and probability (tutorial), Multivariate probability distributions (lecture) and Joint probability distribution (slide).

Once “in” a lecture, a PDF document can be downloaded including all the slides for the lecture.

![Figure 3. Different views into the database of teaching material in the tutor-web](image)

These PDF slides are made with the LaTeX package Beamer (Tantau et al., 2011) and should be ready for classroom use.

A third way of viewing material is on the tutorial level. Users can download a PDF document including all lectures belonging to that tutorial. Each slide is presented as a figure in this document along with all additional material attached to them providing a handout including all relevant information. In a fully developed tutorial this corresponds to a complete textbook.
The length of earthworms in a certain garden follows a normal distribution with mean 11 cm and standard deviation 1.2. If an earthworm is picked at random from the garden what is the probability that it is longer than 12 cm?

\[ \begin{align*}
\text{a)} & \quad 0.7967 \\
\text{b)} & \quad 0.2633 \\
\text{c)} & \quad 0.0333 \\
\rightarrow \text{d)} & \quad 0.2033
\end{align*} \]

We need \( P(X > 12) \) where \( X \sim N(11, 1.2^2) \).

Start by standardizing:

\[
z = \frac{12 - 11}{1.2} = 0.83
\]

We use a normal dist. table and see that for \( z = 0.83 \) we have \( \Phi(z) = 0.7967 \). Remember that \( \Phi(z) = P(Z < z) \).

\[
P(X > 12) = 1 - P(X < 12) = 1 - P(Z < 0.83) = 1 - 0.7967 = 0.2033
\]

R-command: 1-qnorm(12,11,1.2)

*Figure 4. Explanation of the correct answer is given after the student answers a question*

The tutor-web drills are accessible within a lecture. When entering one, a Drill button appears which opens a new tab with the first question when pushed. After answering the question the correct answer is indicated along with some explanation. An example is shown in Figure 4. The question is taken from a basic course in statistics. The material belonging to that course can be found by choosing the Stats department from the welcoming screen and from there Introductory statistics.

**Adding material and content formats**

Teaching material can easily be added to the system through a web-browser. It is important that text-based content as well as mathematical equations and figures are correctly displayed and easily manipulated in a standard browser. To achieve this, several predefined content formats are permitted within the system.

Managers can create departments and courses from the tutor-web homepage. After entering a department teachers can create tutorials that then are linked to one or more courses. Within a tutorial, teachers can subsequently add lectures and later slides. Departments, courses tutorials and lectures are simply collection of slides so they require little more than a name.

After creating a lecture, a tutor-web teacher can create a slide. It can consist of a title and three basic units, the main text, a main figure (graphic) and explanatory text. The format of the main text can be LaTeX (Goossens et al., 1994), plain text, or HTML. The figure(s) can be uploaded files (png, gif or jpeg) or they can be rendered from a text based image format (R-image (R Development Core Team, 2011) or Gnuplot (Williams and Kelley, 2010)). Additional material can be attached to the slides which is available when viewing the material through a browser and in the tutorial printout.

Drill items are grouped together so they correspond to material within a lecture. Questions and answers can be added to the system through a browser or be uploaded from a file. A drill item can have as many answers as desired and there is an option to randomize the order of the answers. The format of the text can be LaTeX or plain text. Questions can therefore include formulas, essential to mathematical instruction. The system also permits the use of the statistical package R (R Development Core Team, 2011) when a question is generated. This allows the generation of similar but not identical data sets and graphs for students to analyse or interpret. Alternatively, a large body of such items can be generated outside the system and then uploaded.

For each item it is possible to put an explanation or solution to the problem along with the question. After a student has submitted his answer the correct answer to the question is displayed along with this explanation.
Case study

It is of particular interest to investigate what affects how students learn in a learning environment, such as the tutor-web. How well a system entices students to continue is a particularly important system feature. One research question is therefore: When do learners decide to stop requesting drill items.

The learners’ responses to drill items in the tutor-web can be used for research on online learning. In the following, data from 316 students in an undergraduate course in calculus will be used. The students were requested to answer drill items from several lectures covering limits, derivatives, logarithms, integration, sequences and series. Within each lecture the students were required to answer a minimum of eight questions correctly but were allowed to continue as long as they liked, with the final eight answers counting towards their grade in the course.

Since all request for items from within a lecture are logged, these appear in a sequence \( n_l = 1,\ldots,m_l \). Data on stopping times can be obtained by looking at the last request of an item from within a lecture, \( m_l \), and we define \( S := I_{nl=ml} \) as a 0/1-indicator variable. One can now formally test which other variables relate significantly to this stopping variable.

The empirical distribution function (edf) of the number of attempts students made within each lecture is given in figure 5. Recall that within this system students are free to make as many attempts as they desire. As a result, the distribution is heavily right-skewed. A jump is seen at 8 attempts: Few students stop before 8 attempts but there is a smooth change in the edf from then on.

![Figure 5. Cumulative distribution (%) of the total number of attempts by each student at each lecture. The right panel expands by only considering attempts 1-25](image)

Finding the drivers

From informal interviews and observations within support sessions that were offered during the semester it seems clear that students have a tendency to continue working within this system until the system reports a high grade. This behaviour is confirmed when looking at the data. Table 1 shows the number of times learners decided to continue requesting drills or decided to stop as a function of the number of correct answers to the last eight items requested within each lecture (the “lecture grade”). As discussed before, only the last eight responses are used to calculate the grade in every lecture and by far, the highest probability (73.3%) is of stopping at the stage when the student has received a full mark (8 out of 8).
Consider next the last response before stopping. Table 2 classifies all responses into groups depending on whether this was the final answer within a lecture and whether the answer was correct or not.

Table 2. Classification of answers according to whether the last question was answered correctly (1) or not (0) and whether the student continued or stopped

<table>
<thead>
<tr>
<th>Last answer</th>
<th>Continue</th>
<th>Stop</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>24665</td>
<td>852</td>
<td>25517</td>
</tr>
<tr>
<td>1</td>
<td>59934</td>
<td>7822</td>
<td>67756</td>
</tr>
<tr>
<td>Total</td>
<td>84599</td>
<td>8674</td>
<td>93273</td>
</tr>
</tbody>
</table>

Naturally, most cases are in the middle of a sequence so most of the observations fall into the “Continue” column. Only about 10% of terminations follow an incorrect response: Unless an incorrect answer follows a long sequence of correct response, it will be beneficial for the student to request another item if the current answer is incorrect.

It has been suggested in earlier studies with this sort of grading scheme as used here that students may decide to stop early if a run of correct responses is followed by an incorrect answer (Stefansson and Sigurdardottir, 2011). To investigate this, one can consider the fraction of stopping as a function of both the current lecture grade and the most recent grade. This is shown in table 3.

Table 3. Fraction of stopping (%) as a function whether the last questions was answered correctly (0) or not (1) and the number of correct answers in the last eight questions

<table>
<thead>
<tr>
<th>Last = 0</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.9</td>
<td>1.5</td>
<td>1.3</td>
<td>0.8</td>
<td>1.0</td>
<td>2.4</td>
<td>5.4</td>
<td>24.7</td>
</tr>
<tr>
<td>Last = 1</td>
<td>-</td>
<td>2.4</td>
<td>1.4</td>
<td>1.4</td>
<td>2.1</td>
<td>2.0</td>
<td>3.9</td>
<td>8.0</td>
</tr>
</tbody>
</table>

It is seen in table 3 that if a run of 7 correct answers is followed by an incorrect answer then the student will in 25% of all cases decide to stop. This is a perfectly logical result since a student who has a sequence of 7 correct and one incorrect, will need another 8 correct answers in sequence to increase the grade. An improved averaging scheme can be used to alleviate this problem. An algorithm which uses the average of the most recent \( \min(n/2, 30) \) grades after \( n \) attempts can give the same full grade after 8 correct responses but a single incorrect answer will get much lower weight as more correct answers come in. An even better option could be to use tapering to downgrade old responses. Given the incentive to work towards a high grade (table 1), this simple change is likely to alleviate the 25% stopping problem.

In principle a generalized linear model (assuming a binomial distribution and logit link) can be fitted to the data shown in table 1. As can be seen in the table, the relationship between the probability of stopping and the grade is not a linear one. A factor variable with four levels (low grade = 0–2, median grade = 3–5, high grade = 6–7 and top grade = 8) will be used. The results are shown in table 4. Here the 0/1 indicator of whether the student stopped is “regressed” against the factor variable grade at that timepoint. The statistical package R (R Development Core Team, 2011) was used for the analysis.

Although this is a useful approach to estimating effects and obtaining an indication of p-values, assumptions can not be assumed to be completely correct since there will be a subject effect (this is considered below). Nevertheless the estimates indicate that the probability of stopping is increasing with higher grade. The difference in probability of
stopping between the low grade (0–2) and the median grade (3–5) is not significant. The difference between the high-grade and the low grade as well as the top-grade and low grade are highly significant.

Table 4. Parameter estimates where the 0/1 indicator of whether the student stopped is regressed against the grade at that timepoint

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. error</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-4.2901</td>
<td>0.1591</td>
<td>-26.97</td>
</tr>
<tr>
<td>Median grade</td>
<td>0.2733</td>
<td>0.1634</td>
<td>1.67</td>
</tr>
<tr>
<td>High grade</td>
<td>1.6002</td>
<td>0.1613</td>
<td>9.98</td>
</tr>
<tr>
<td>Top grade</td>
<td>5.3018</td>
<td>0.1613</td>
<td>32.86</td>
</tr>
</tbody>
</table>

In order to see which other variables are related to the probability of stopping a more complicated model, also including an indicator variable stating if the last answer was correct or not (grade_last), difficulty of the question (diffic) (computed based on the proportion of incorrect responses) as well as the number of attempts (nattl) was fitted to the data. One can see in Table 5 that these all appear highly significant. Also notice the sign of the parameter estimates: the probability of stopping increases with higher grades but decreases with increased difficulty and number of questions answered. Also, the probability of stopping decreases if the last question was answered correctly in comparison to when the last question was answered incorrectly.

Table 5. Parameter estimates where the 0/1 indicator of whether the student stopped is regressed against the grade at that timepoint, grade of last answer, item difficulty and number of items answered

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. error</th>
<th>z-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-3.7509</td>
<td>0.1645</td>
<td>-22.80</td>
</tr>
<tr>
<td>median grade</td>
<td>0.3764</td>
<td>0.1642</td>
<td>2.29</td>
</tr>
<tr>
<td>high grade</td>
<td>1.7813</td>
<td>0.1631</td>
<td>10.92</td>
</tr>
<tr>
<td>top grade</td>
<td>5.4785</td>
<td>0.1656</td>
<td>33.07</td>
</tr>
<tr>
<td>grade_last</td>
<td>-0.2801</td>
<td>0.0444</td>
<td>-6.30</td>
</tr>
<tr>
<td>diffic</td>
<td>-0.2946</td>
<td>0.1354</td>
<td>-2.18</td>
</tr>
<tr>
<td>nattl</td>
<td>-0.0182</td>
<td>0.0010</td>
<td>-18.51</td>
</tr>
</tbody>
</table>

When estimating an even more complicated model, also including the student effect and the lecture (or content) effect it was found that both factors were highly significant.

Conclusions and future work

These simple analyses clearly indicate how one can potentially improve upon the item database and the grading scheme. Given the strong incentive for students to obtain high grades in this kind of system, it is imperative that the system includes a wide variety of very difficult problems, since clearly students continue well into the most difficult parts of the content. It is also clear from this analysis that the behaviour of the students is affected by the grading scheme used here. The effect of different grading schemes on students’ behaviour are therefore currently under investigation. Instead of using the last eight responses and giving them equal weight an experiment with tapering schemes is being designed where the most recent answers get more weight than older answers. In these experiments the students will be randomly assigned a value that determines how much weight is put on their most recent answer. Before requesting a new drill item the students are told what their grade will be if they answer the next drill item correctly, with the intent of enticing them to continue requesting drill items and learning more. When large weight is given to the most recent answer and the student answers the item correctly, the student receives an immediate award (large jump up in grade). However if the answer is incorrect the student gets immediate punishment (large jump down in grade). The risk is that the student then stops requesting items but the intent is that information on the potential grade increase will tempt students to continue working within the system.

Currently the tutor-web content provider sets up a concept map by structuring the teaching material within a course into tutorials and lectures within tutorials, providing a linear and logical learning path with respect to prerequisites. Drill items are grouped within lectures and are chosen according to the learners ability resulting in individualized
path of drill items for each learner within a specific topic. One of the goals of the tutor-web project is to link appropriate learning material to the drill items so if a student answers an item incorrectly the student will be pointed towards appropriate material to read. These links will be made to material within the system but it would also be interesting to allow users to provide links to other Creative Commons licensed material outside of the system resulting in a completely individualized learning path through an entire course within the system or even the entire web.

The tutor-web is an ongoing research project into the online student’s behaviour. An experiment was made in an older version of the system to assess potential difference in student learning while working in the tutor-web versus students handing in written homework (Jonsdottir and Stefansson, 2011). The difference in learning between the groups was not found to be significant but more importantly the confidence bound for the difference was tight, indicating no difference of importance. This implies that time spent on correcting written assignments can be saved by using the tutor-web as homework instead of some written assignments, potentially making a considerable financial difference. The system is under constant development and these results imply that further improvements in learning through the system will enhance it to become better than traditional assignments. Current research is therefore focused on ways to amplify learning rather than changing from one medium to another.

An interesting research question is: How should items be allocated so the students get the most out of the drills? Current CAT techniques tend to use Point Fisher Information (PFI), justified when attempting to evaluate current knowledge since with the PFI the selection criterion is to minimize the variance in estimated subject ability after seeing the item. With the current emphasis on learning, not evaluation, the PFI is no longer central. Although one can in principle still use the PFI methodology, the basic criteria for using the PFI are no longer of interest: Instead, one wants to select each item so as to maximize the amount of learning obtained by showing the item to the subject. In addition, one wants to make sure that this learning is not just transient but committed to long-term memory and, if, at all possible that learning occurs with understanding - not simply learning by rote. Such a mechanism for selecting items could and should take a number of concerns into account.

- If selecting within the current lecture, select an item to give maximum learning
- Within the lecture select easy items for an underachiever, hard items for a good student
- Increase the difficulty level as the student learns, within the lecture
- Select items so that a student can only “successfully” complete the lecture by completing the most difficult items
- Select items from previous lectures (or prerequisite tutorials/courses) if the student cannot handle the easiest items within the lecture
- Estimate whether the student is likely to be forgetting earlier material and select earlier items accordingly
- Select an item based on externally supplied metadata on item taxonomy, such as an item containing a cue
- Select items from material which the student has earlier answered incorrectly or is likely to answer incorrectly

Some of these points are already implemented as described in this paper but others form future research projects.

The system has been used by over 2000 students, mostly in courses on statistics and mathematics at the University of Iceland (UI) and the University of Maseno, Kenya. The tutor-web has mainly been used to supplement education in the classroom, but being freely accessible without regard to physical location or registration into a school or university, the potential is much greater. Completion of certain courses has e.g., been used as an entry criterion for PhD applicants at the UI, as an addition to other formal criteria for entering a PhD study. Similarly the system can be used by students lacking in prerequisites.

The tutor-web has considerable potential for low-income regions like Kenya where textbooks are not widely available, and student surveys regarding the tutor-web are highly positive, “I wished to do more” being a typical response (Mokua, Stern, Jonsdottir & Mbasu, 2013). A mobile tutor-web is under development, where the user does not need to be connected to the Internet at all times to answer drill items. This can become a game changer for students in rural areas where Internet access is limited but the number of students with access to smart phones is exploding. Hopefully increase the number of students whose experience will be described by words from the Maseno survey: “Doing maths online was the best experience I ever had with maths.”
Acknowledgements

The tutor-web project has been supported by Ministry of Education and the Marine Research Institute of Iceland, the United Nations University, University of Iceland and the EU project MareFrame. The authors would finally like to thank Christopher Desjardins for his helpful comments.

References


Integrating Learning Services in the Cloud: An Approach that Benefits Both Systems and Learning

Gustavo Gutiérrez-Carreón1,3∗, Thanasis Daradoumis1,2 and Josep Jorba1

1Open University of Catalonia, Av. Tibidabo 39-43 - 08035 Barcelona, Spain // 2University of the Aegean, University Hill, 81100, Mytilene, Greece // 3Universidad Michoacana de San Nicolas de Hidalgo, Francisco J. Mújica S/N, Morelia, Michoacan, Mexico // g Gutierrezc@uoc.edu // adaradoumis@uoc.edu // jjorbae@uoc.edu

∗Corresponding author

ABSTRACT

Currently there is an increasing trend to implement functionalities that allow for the development of applications based on Cloud computing. In education there are high expectations for Learning Management Systems since they can be powerful tools to foster more effective collaboration within a virtual classroom. Tools can also be integrated with these systems that can provide assistance to students on any topic covered in class through the proliferation and popularity of publicly available Web APIs. However, generally most students and teachers still face technological barriers to access these resources and integrate them into their educational systems. This paper discusses an effort to overcome this problem by proposing a method of resource access through linked data, which is explicitly defined with semantic annotations related to the learning environment. Doing so, we are able to decorate the Cloud providers API’s, thus providing easy integration into learning applications and important benefits for learning. The effectiveness of our approach has been measured and evaluated through usability methods and cognitive load theory principles.

Keywords

Cloud computing, Semantic web, Learning service, Usability, Cognitive load

Introduction

Cloud computing is a distributed computing paradigm under several service models (Linthicum, 2009). The Software as a Service model (SaaS) is a software and data related delivery model, where data is centrally hosted on the Cloud as a service. Under a public Cloud deployment (Baun et al., 2011), the services are offered by subscription and accessed over the Internet. The private Cloud (Rosenberg & Mateos, 2011) enables organizations to build and maintain data centers at low costs, implement new systems quickly and easily, provide resources elastically and reducing the need for organizations to maintain a large IT staff or infrastructure. In education, Cloud computing technology (and the construction of platforms for college education management) can improve the utilization resource rate, saving university resources and improving teaching level. It can also bring new areas of application closer to real life problems and study areas (Dong et al., 2012). Each type of Cloud has some kind of Application Program Interface (API), called Cloud Service API (Morel et al., 2011), that can be used to provide resources, configure, control and release them when no longer needed.

Vaquero (2011) presents an evaluation of the real benefits of Cloud computing for a course on networks using Cloud Paas (Platform as a Service) and IaaS (Infrastructure as a Service). The results show that the introduction of Cloud technology is adequate to maintain students' attention and save time in the tasks related to the use of technologies to support education. However the same results show that the use of Cloud technology and levels of abstraction "per se" are not enough to help motivate students neither provide any considerable improvement as regards the results obtained or the acquisition of new knowledge. In this sense, the use of Cloud technology cannot be oblivious to learning objectives. Instead, it must be integrated with certain educational methodologies so that it makes a real difference in student-centered learning activities. For instance, this study showed that the use of technology "per se" does not generate the expected benefits in students’ outcomes, but when this technology is used to foster activities that include collaboration, interaction, and monitoring of student progress, amongst others, the outcomes are better. In this regard, Manro et al. (2011) highlight how Cloud computing can be used to solve educational problems. This is especially true for tools that use the SaaS model, such as Facebook and Twitter or SaaS providers such as Google Apps, which is the Cloud scheme that most universities are already implementing. Similarly, this study mentioned some general benefits that the use of Cloud tools may have on university-level learning, but it has not discussed how these tools can be used in conjunction with e-learning applications or LMS and, most importantly, whether this integration may have specific benefits to both learning and teaching.
Given the prevalence of Cloud-based services in students' and faculty's personal lives, it is without a doubt necessary to consider ways in which these can be combined for the purposes of teaching and learning. It is therefore relevant to investigate how Cloud-based services can be used for pedagogy.

Our work takes the promise of Cloud and their models like SaaS a step further, which substantially benefits the development of learning systems. This is made possible by simplifying the use of the functionality provided by each Cloud Service API and integrating it to the educational environment in an easier and more flexible way. To achieve it, this paper proposes a semantic mechanism for integrating a Cloud service API with an educational system. In fact, we focus on issues related to usability (Hollender et al., 2010) and Cognitive Load Theory – CLT (Plass et al., 2010) that have to be taken into account in a holistic way. This approach is followed in order to determine whether our proposed solution for integrating learning services in the Cloud can benefit both systems and learning. On the one hand, it is the basic assertion of CLT that any instructional design needs to take the limitations of working memory into account in order to prevent an overload of working memory capacity and hence a deterioration of learning. On the other hand, usability is defined as the extent to which a user can fulfill a task using a tool effectively, efficiently, and with satisfaction; moreover, the level of the usability of a tool or application can be defined only in the context of its specific users and the specific tasks that are to be accomplished. To get to this point, first we discuss related work and then we present an approach that discusses how technologies can support semantic web and Cloud computing-based services. Subsequently, a case study related with the integration of Google Apps with Chamilo (Maes, 2010) Learning Management System (LMS) through semantic description of services is presented. Then, we discuss the results of its implementation. We conclude with suggestions for future research.

**Related work**

There are some related efforts to the design and development of learning systems that take advantage of the features and benefits provided by Cloud computing.

First, virtualization efforts try to create learning environments more cost effectively without compromising the level of service or user experience. Such an example is presented by Hu et al. (2011), where a successful change of the use and payment patterns for a Web-based virtual learning environment and pedagogical software was achieved through virtualization and SaaS enablement of web services. There are also efforts that demonstrate the efficiency and effectiveness of services collaboration in the Cloud. This is the case of AMBAR-C (Awareness-based learning Model for distributive collaborative enviRonment) (Paletta and Herrero, 2010), a collaboration model used for a multi-agent based system in collaborative Cloud computing environments. In the area of micro-learning applications development, mobile services have a great potential in supporting informal learning processes. Kovachev et al. (2011), propose a tool that makes use of Cloud services to promote ubiquitous learning.

All these learning systems try to take full advantage of Cloud computing for the development of learning systems, but we see that the integration of existing learning tools with functionalities available from Cloud providers presents several problems and limitations, especially those based on SaaS. Such integration is important because it facilitates the definition of Cloud services’ APIs. The next section presents the main objective and research questions of this study. It is followed by a description of the approach and discusses how technologies can support the integration of e-learning systems and Cloud services when partly based on the Semantic Web.

**Objectives and contribution**

Despite the efforts made in related research work, it is not yet clear how the integration of Cloud computing-based services can benefit both e-learning systems and students’ learning processes. Consequently, the aim of this study it to explore the workability of a semantic description method which, on the one hand, defines the characteristics of learning services using terms that are more closely related to the educational field, thus facilitating the integration of Chamilo LMS with Cloud services. On the other hand, it wants to ensure benefits for students in terms of usability issues and cognitive overhead/load. To achieve the latter objective and test the effectiveness of the proposed approach, we set the following research questions:
• Is there a significant difference between the cognitive load-satisfaction of the students who learn with an LMS integrated with Cloud services and the students that manually access to both systems separately?
• Is there an acceptable System Usability Scale (SUS) from students who learn following the LMS integrated with Cloud services compared to those students that use independent systems?

Our study showed that the integrated system, in which Cloud services and the LMS are used as one seamless system, is more desirable and user-friendly, whereas it proved to be less strenuous on students’ cognitive loads.

An approach to integrate Cloud services with e-learning systems across linked data

With an ever-increasing list of services that are provided through the Cloud, many critical applications will be deployed and consumed through SaaS mechanisms in the near future. However, as with any new technology, the concepts of SaaS and the Cloud have limitations and problems, especially those related with the integration of applications and data sources. Integration has been an important subject of study and research which seek to determine how integration brings a sense of order out of the chaos and disorder created by heterogeneous systems, networks and services (Raj, 2011).

The service-oriented architecture (SOA) was conceived as an architectural meta-model at a higher abstraction level to the technologies that are used to implement it. SOA is currently implemented using Web service technologies such as Web Service Description Language (WSDL) and Simple Object Access Protocol (SOAP). WSDL was specifically created for this purpose, to provide several constructions which are useful to services through an XML schema, including those that describe the service operations and methods, the parameter descriptions as well as the information about the type of Protocol which is necessary to invoke the service. Figure 1 illustrates the basic interactions between the service provider, services requester and services directory for dynamic binding of services at runtime (Baun et al., 2011). In particular, a service requester (consumer) can locate a suitable service in a service directory. If a suitable service has been found, the service consumer receives a reference for accessing the service. Then, the service can be called. The service provider replies by sending a message back. Today more and more APIs are being published by the major players in the Web, such as eBay, Amazon and Google. Web applications that consume these APIs are collectively referred as mashups, and they offer interesting experience to web users.

![Service Management through WSDL](image)

Figure 1. Service management through WSDL

However, many activities require additional features which are not caught in the specification of basic services supported by WSDL. Consequently, the so called semantic web service standards and composition languages have been defined to provide additional capabilities to service-oriented solutions that require discovery, composition, orchestration, choreography and mapping of web services. The basics of representing the Semantic Web consist of the use of two core standards: the Resource Description Framework (RDF) (Brickley & Guha, 2004) and the Web Ontology Language (OWL) (Allemang & Hendler, 2011). Additionally, linked data (Hogan et al., 2012) refers to data published on the web in such a way that it is machine readable, its meaning is explicitly defined, it is linked to other external datasets, and it can in turn be linked to from external datasets as well.

147
In the particular case of learning systems, integration problems are associated with the different understandings that may hold in the educational field and which are related to the particular characteristics of each area of instruction. The functional characteristics of a Cloud service API often do not match the needs that students and teachers may have. For example, there may be an API that allows users to handle forms and process them, but in the educational field a teacher would try to access this functionality with the aim of providing “evaluation and feedback” to students. In that sense, we propose an alternative approach that makes use of the Semantic Web in general and the semantic description of services in particular to enrich the functional characteristics of the services or API’s with ontological domain characteristics of each area of teaching.

More specifically, given a specific teaching area, we propose to employ semantic annotations that will enable us to use the API functionality according to the ontological model used in the corresponding learning system, to let “decorate” and facilitate their integration with other tools. In this case, search is also promoted through characteristics that are common for teachers and students. Figure 2 shows a graphical representation of the semantics of a learning service, which is the result of a Cloud service API plus the semantic attributes that help us operate it and its OWL class representation.

![Figure 2. A semantically represented learning service and its ontology in OWL](image)

This fact does not affect or interferes with the processes of each service. In the case of search, several features facilitate the human interaction with the service, such as: a service name, a readable description, a service category, a type of collaboration that is designed for the service, a type of connection devices which can display the result of the service, and the user profiles that can access the service. As shown in Figure 3, RDF storage is fed with instances of OWL class with defined parameters contained in the description of Cloud services APIs. Subsequently these containers are exploited by mashups to get the services that meet the requirements of users. These results contain functional specifications for invoking them.

Comparing our approach with other studies, as the ones mentioned in the previous two sections, the integration of learning tools with functional Cloud services across semantic models represents significant benefits for the design of learning systems since it establishes common data containers for both applications, far from specific and rigid data models where each is defined separately. The educational impact that represents this type of innovation is that teachers are able to work with an integrated environment which facilitates the interaction between applications and thus provides them better opportunities in developing more solid and useful learning activities. Moreover, students
achieve to reduce cognitive load (especially on task performance and the time to complete them) and work in a more user-friendly environment, as we demonstrate in the following sections.

![Figure 3. RDF storage fed with OWL instances and exploited with mashups](image)

**Experiment design**

We have designed an experiment which deals with the integration of Google Apps Cloud features (Morel et al., 2011) within the Chamilo learning management system. In general, Chamilo supports many different kinds of learning and collaborative activities. It allows one to specify the course objectives, identify learning units, develop materials and presentations for content, build evaluation exercises and on-line tasks, while it provides assistance and collaboration tools that contain synchronous and asynchronous communication (chat and forums), as well as wikis, blogs, social networks, messages, group management, course management, recycling and more. However, instructors may have specific needs in order to meet particular objectives of a course, for example they may need a tool that is beyond the scope of Chamilo, such as a diagram designer, mail services, a document processor, etc. Such tools are offered by Google Apps and include Google Drive for storage and document creation, Youtube for audiovisual content management, Google Hangouts for handling instant electronic messaging, and live video conferencing capabilities, amongst others.

The main aim of the experiment is to develop semantic mashups to integrate the two systems using Linked data and taking advantage of the availability of APIs to interact with Google Apps.

**Participants**

The experiment designed above has been carried out in a class of third-year undergraduate students ($N = 56$) enrolled in the “Management Information Systems” course. This course is of both theoretical and practical type, and the teacher had already taught it in two previous occasions. All students have been using Chamilo as a standard tool to access course documents, submit assignments, attend meetings, and so forth. The students participated in the experience were divided into two groups: an experimental group ($N = 24$) and a control group ($N = 32$). Students in the experimental group used Chamilo integrated with Cloud services through semantic mashups and Linked data,
whereas the control group was asked to access both tools separately, performing the Cloud services access manually. Both groups of students were instructed by the same teacher. The teacher assigned students specific problems to solve using the project-based learning paradigm and guided students to use Google Drive and Google Docs to share and co-construct documents, Google Groups to carry out structured discussions and Google Calendar to organize task planning and execution. The teacher provided the experimental group students with very specific indications of how and when to use each application, along with Chamilo, which definitely had a positive impact on cognitive load and usability.

Experimental procedure

1. Three 90-minute sessions over a period of two weeks were conducted, totaling 270 minutes, where all students had the chance to give a general explanation that showed how to access and use Google Apps main applications, such as: Google Drive (including Docs), Gmail, Translate, Calendar, YouTube, Sites, Hangouts, and Maps. Both experimental and control groups used Chamilo for daily course activities.

2. The learning experiment, that lasted 120 minutes, consisted of three activities that made use of Google Apps:
   - Generate a site to publish information on an assigned topic, using Google Sites.
   - Design 2 graphics to incorporate into each personal site using Google Docs.
   - Incorporate a video explaining the process of designing graphics in each personal site using YouTube.

Finally the site had to be linked to the profile of each student in the Chamilo platform. Students in the experimental group were guided to work using the semantics mashups integrated with Chamilo, while students in the control group worked on their experiment accessing Cloud services manually.

After the end of the learning experiment, all students took a usability questionnaire as well as a questionnaire that examined students’ cognitive-load satisfaction.

Evaluation instruments

As commented above, one of the objectives of this study is to determine the benefits that students can get from Cloud service integration with LMS. These benefits may occur at two levels: first, at cognitive load level, we examine whether students that have access to the integrated functionality have less cognitive overhead and thus focus their attention better on learning and, second, at usability level. Usability is defined as the extent to which a user can fulfill a task using a tool effectively, efficiently, and with satisfaction (ISO 9241, 1998). In fact, these two goals are interrelated, since the use of a complex educational environment that takes standard usability guidelines and principles into account may contribute to reduce extraneous cognitive load (Hollender et al. 2010). Furthermore, the applicability of existing cognitive-load educational design principles for educational software design should be evaluated empirically. It should not be assumed that Cognitive Load Theory (CLT) and its instructional design principles offer off-the-shelf solutions for educational technology.

The design and use of learning tools for online education must be done carefully when dealing with CLT issues; for instance, Wong et al (2012) found that if we assume that the cognitive load associated with the presentation of complex information could be improved by the use of animations or dual mode presentations, both forms of presentation incidentally introduce transience that also can impose a heavy cognitive load. Despite these shortcomings, there have been successful experiments, in which the consideration and evaluation of cognitive load has represented benefits to the development of learning scenarios (Schrader & Bastiaens, 2012; Moons & De Backer, 2013).

To design an instrument that allows us to evaluate and take cognitive load into account, we rely on a proposal developed by Bradford (2011), who detected a coefficient for the relationship between satisfaction and cognitive load by separating academic performance (i.e., “learning”) from cognitive load and satisfaction. In this sense a survey instrument was designed, which focuses on three indicators to consider: awareness, challenge and engagement.
A survey of 10 items was implemented with a 5-point Likert scale. As concerns the awareness indicator, four questions were posed to examine whether syllabus and assignment directions were clear, problems were easy to comprehend, supporting materials were helpful, and activities were useful for new cases (questions 1 - 4). The fact of being aware of all these elements constitutes important means for finding solutions and organizing presentations so as to reduce high memory load and enhance student motivation. With regards to the challenge indicator, three questions relate the students’ degree of satisfaction with the degree of challenge they face (questions 5 - 7); it is expected that when students are challenged, satisfaction may be their own reward and extra memory requirements (i.e., high cognitive load) seem to be fine. As for the engagement indicator, three questions were asked to relate relevance of different types of learning activities to the students’ needs and goals (questions 8 - 10). Figure 4 presents the 10 questions used in the cognitive load - satisfaction questionnaire. Here we note that although a weak reliability (.49) exists, the coefficient can be provisionally accepted as a new scale, given that follow-up efforts to improve the coefficient are made.

![Figure 4. Questionnaire that examines the relationship of cognitive load – satisfaction](image)

With regard to the second goal, namely the evaluation of usability issues, we asked students to respond to a usability questionnaire, called System Usability Scale (SUS). SUS is a simple, ten-item attitude 5 point Likert scale (ranging from 1-'strongly disagree' to 5 - 'strongly agree'), giving a global view of subjective assessments of usability. It was developed by Brooke (1996) and its validity has been tested by numerous studies both in several websites and Learning Management Systems as well as in other environments, such as mobile ones. It proved to produce very reliable outcomes in relation to other questionnaires, even for a small ($N = 12-15$) sample of participants (Tullis and Stetson 2004). It yields a single score on a scale of 0–100. Extensive studies with the participation of almost 2300 users confirmed that the mean evaluation value has been 70, whereas the top 25% of all scores was measured at 77.8 (Bangor et al. 2008; Bangor et al. 2009). In particular, Bangor et al. (2008) showed the following qualitative interpretation of SUS scores:
- $SUS = 51 \Rightarrow$ Poor/OK
- $SUS = 72 \Rightarrow$ Acceptable/Good
- $SUS = 85 \Rightarrow$ Excellent

Figure 5 shows this in more detail.

<table>
<thead>
<tr>
<th>Adjective</th>
<th>Count</th>
<th>Mean SUS Score</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Worst Imaginable</td>
<td>4</td>
<td>12.5</td>
<td>13.1</td>
</tr>
<tr>
<td>Awful</td>
<td>22</td>
<td>20.3</td>
<td>11.3</td>
</tr>
<tr>
<td>Poor</td>
<td>72</td>
<td>35.7</td>
<td>12.6</td>
</tr>
<tr>
<td>OK</td>
<td>211</td>
<td>50.9</td>
<td>13.8</td>
</tr>
<tr>
<td>Good</td>
<td>345</td>
<td>71.4</td>
<td>11.6</td>
</tr>
<tr>
<td>Excellent</td>
<td>289</td>
<td>85.5</td>
<td>10.4</td>
</tr>
<tr>
<td>Best Imaginable</td>
<td>16</td>
<td>90.9</td>
<td>13.4</td>
</tr>
</tbody>
</table>

Figure 5. SUS Adjective rating - by Bangor et al. (2009)

Figure 6. The System Usability Scale (SUS) questionnaire
Finally, Tullis and Albert (2008), after conducting 129 studies, concluded that a score greater than 81.2 implies ranking at the top 10%. They also found that a score greater than 80 implies an increased likelihood of returning to the website and recommending it to a friend or acquaintance.

Figure 6 presents the 10 questions used in the SUS questionnaire. From the 10 questions, 5 are positively-worded and 5 are negatively-worded, alternating each positive with a negative question; by alternating positive and negative items, the respondent has to read each statement and make an effort to think whether they agree or disagree with it.

To calculate the SUS score, we first sum the score contributions from each item. Each item's score contribution will range from 0 to 4. So, for items 1,3,5,7 and 9 the score contribution is the scale position minus 1. For items 2,4,6,8 and 10, the contribution is 5 minus the scale position. Finally, we multiply the sum of the scores by 2.5 to obtain the overall value of SUS in the range of 0 to 100.

**Presentation and significance of the results**

After students completed the experiment, they were asked to answer the two questionnaires presented above. The first questionnaire assessed the relationship between satisfaction and cognitive load that students experienced while trying to perform the activities of the experiment. The second questionnaire evaluated the students’ experience in terms of usability of the tool. The descriptive statistical analysis includes the mean, median, mode, standard deviation, variance, skewness, kurtosis, range, maximum, minimum, and sum of the response to each question. These statistics provide an overview of the students’ responses to the questionnaire. The values of the responses are those of a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree), so they are discrete values used to provide basic statistical calculations such as the mean, standard deviation, variance, skewness and kurtosis, which reflect the trend of values of the responses and the shape of a probability distribution, without representing themselves a specific value within the scale. The case of values, such as mode, median, minimum value, maximum value, sum and range, shows the values of the options in which students responded, giving us a breakdown of the options selected in each response.

The results of the first questionnaire are presented in Tables 1 and 2 below, which reflect a statistical profile of the students’ answers to each of the questions for both the control and the experimental group. In general, in each of the evaluation criteria, the results of the experimental group were closer to strongly agree compared to the same results of the control group. In particular, to determine the relationship between cognitive load and satisfaction, we examined the three indicators defined above: awareness, challenge and engagement.

<table>
<thead>
<tr>
<th>Table 1. Descriptive statistics estimated for the control group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>SE of the mean</td>
</tr>
<tr>
<td>Median</td>
</tr>
<tr>
<td>Mode</td>
</tr>
<tr>
<td>Standard deviation</td>
</tr>
<tr>
<td>Variance</td>
</tr>
<tr>
<td>Skewness</td>
</tr>
<tr>
<td>Kurtosis</td>
</tr>
<tr>
<td>Range</td>
</tr>
<tr>
<td>Minimum</td>
</tr>
<tr>
<td>Maximum</td>
</tr>
<tr>
<td>Sum</td>
</tr>
</tbody>
</table>

In the case of the awareness indicator (questions 1 - 4), based on the responses of each of the groups, we can determine that the experimental group showed a better performance for finding solutions and organization of the presentation. We can conclude that, compared with the control group, the integrated system reduced high memory load and increased the motivation of students. As for the challenge indicator (questions 5 - 7), the experimental
group showed a better relationship between the degree of student satisfaction and the level of challenge they face; in this case satisfaction has been students’ main reward, which eases the additional memory load required, thus alleviating cognitive load and increasing students’ general well-being. As concerns the engagement indicator (questions 8-10), the experimental group showed a better perception of the relevance of different types of learning activities with regard to the students’ needs and goals, which improves the ‘cognitive load – satisfaction’ relationship for these students. All in all, we can conclude that students using the integrated system had significant benefits as regards the ‘cognitive load- satisfaction’ relationship with respect to the control group students that used Google Apps services independently.

Table 2. Descriptive statistics estimated for the experimental group

<table>
<thead>
<tr>
<th>Question</th>
<th>N</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
<th>Q6</th>
<th>Q7</th>
<th>Q8</th>
<th>Q9</th>
<th>Q10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>SE of the mean</td>
<td></td>
<td>.1009</td>
<td>.1009</td>
<td>.1241</td>
<td>9.8E-02</td>
<td>.1320</td>
<td>.1009</td>
<td>.1039</td>
<td>.1009</td>
<td>.1009</td>
<td>9.4E-02</td>
</tr>
<tr>
<td>Median</td>
<td></td>
<td>5.0000</td>
<td>5.0000</td>
<td>4.0000</td>
<td>5.0000</td>
<td>4.0000</td>
<td>5.0000</td>
<td>5.0000</td>
<td>5.0000</td>
<td>5.0000</td>
<td>5.0000</td>
</tr>
<tr>
<td>Mode</td>
<td></td>
<td>5.00</td>
<td>5.00</td>
<td>4.00</td>
<td>5.00</td>
<td>4.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
</tr>
<tr>
<td>Standard deviation</td>
<td></td>
<td>.4945</td>
<td>.4945</td>
<td>.6079</td>
<td>.4815</td>
<td>.6469</td>
<td>.4945</td>
<td>.5090</td>
<td>.4945</td>
<td>.4945</td>
<td>.4643</td>
</tr>
<tr>
<td>Variance</td>
<td></td>
<td>.2446</td>
<td>.2446</td>
<td>.3696</td>
<td>.2319</td>
<td>.4185</td>
<td>.2446</td>
<td>.2591</td>
<td>.2446</td>
<td>.2446</td>
<td>.2156</td>
</tr>
<tr>
<td>Skewness</td>
<td></td>
<td>-.551</td>
<td>-.551</td>
<td>-.158</td>
<td>-.755</td>
<td>-.542</td>
<td>-.551</td>
<td>-.179</td>
<td>-.551</td>
<td>-.551</td>
<td>-.979</td>
</tr>
<tr>
<td>Kurtosis</td>
<td></td>
<td>-1.859</td>
<td>-1.859</td>
<td>-.347</td>
<td>-1.568</td>
<td>-.519</td>
<td>-1.859</td>
<td>-2.156</td>
<td>-1.859</td>
<td>-1.859</td>
<td>-1.145</td>
</tr>
<tr>
<td>Range</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td>2.00</td>
<td>1.00</td>
<td>2.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Minimum</td>
<td></td>
<td>4.00</td>
<td>4.00</td>
<td>3.00</td>
<td>4.00</td>
<td>3.00</td>
<td>4.00</td>
<td>4.00</td>
<td>4.00</td>
<td>4.00</td>
<td>4.00</td>
</tr>
<tr>
<td>Maximum</td>
<td></td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
<td>5.00</td>
</tr>
<tr>
<td>Sum</td>
<td></td>
<td>111.00</td>
<td>111.00</td>
<td>102.00</td>
<td>112.00</td>
<td>105.00</td>
<td>111.00</td>
<td>109.00</td>
<td>111.00</td>
<td>111.00</td>
<td>113.00</td>
</tr>
</tbody>
</table>

In addition to the questionnaire responses of Tables 1 and 2 that correspond to a Likert scale with discrete response values ranging from 1 to 5, we considered useful to include frequency tables which show the number (and percentage) of students who answered each of the options (1 strongly disagree, 2 disagree, 3 neutral, 4 agree and 5 strongly agree) associated with each question. Tables 3 and 4 show the frequency tables for the responses of each question for the control and experimental groups respectively.

Table 3 shows that the results of the control group fluctuate between the values of 2 representing disagree, to 5 representing totally agree, and are divided into thirds responses between options 3, 4 and 5. In the case of the experimental group, Table 4 shows that values range from 3- neutral to 5- strongly agree and the answers are divided into halves on value 4 corresponding to agree and value 5 corresponding to strongly agree. From these results we can highlight some points that indicate that the LMS tool integrated with Cloud services is more useful and effective for the students in the experimental group and, in fact, improves the cognitive load-satisfaction relationship. As a general comment, we can see that in the majority of questions (8 out of 10), all the students (100%) of the experimental group have a positive attitude (i.e., agree or strongly agree). There are only 2 questions (Q3 and Q5) with a few (8.3%) neutral responses. In contrast, students in the control group were not so positive. Even in some instances they expressed disagreement. For example, while in Q1 100% of the experimental group students agree or strongly agree that the instructions and guidelines were clear, only 50% do so in the control group, with 15.6% of students showing disagreement. Similarly, in Q4 only 37.5% of the control group students strongly agree that the knowledge acquired in the experiment will be useful in future projects, against 66.7% in the experimental group, which clearly shows the higher level of satisfaction of the experimental group students. Finally, in Q10 the experimental group shows more confidence to develop these kinds of projects with 70.8% strongly agree against only 28.1% of the control group.

Finally, we performed the reliability measure of the Cognitive Load instrument based on Cronbach’s alpha which considers as acceptable value a coefficient around 0.8. Table 5 shows the results for the binding coefficient of the experimental and control groups, as well as their independent values. The results demonstrate that the reliability of the instrument used is quite good in all cases, showing a clear consistency in the case of the experimental group. As a conclusion, the instrument used to address our first research question (i.e., the difference between the cognitive load-satisfaction of the students who learn with an LMS integrated with Cloud services and the students that manually access to both systems separately) proves to be reliable.
Table 3. Frequency table for the control group

<table>
<thead>
<tr>
<th>Likert scale</th>
<th>2</th>
<th>%</th>
<th>3</th>
<th>%</th>
<th>4</th>
<th>%</th>
<th>5</th>
<th>%</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>5</td>
<td>15.6</td>
<td>11</td>
<td>34.4</td>
<td>5</td>
<td>15.6</td>
<td>11</td>
<td>34.4</td>
<td>32</td>
</tr>
<tr>
<td>Q2</td>
<td>1</td>
<td>3.1</td>
<td>10</td>
<td>31.3</td>
<td>9</td>
<td>28.1</td>
<td>12</td>
<td>37.5</td>
<td>32</td>
</tr>
<tr>
<td>Q3</td>
<td>1</td>
<td>3.1</td>
<td>9</td>
<td>28.1</td>
<td>16</td>
<td>50.0</td>
<td>6</td>
<td>18.8</td>
<td>32</td>
</tr>
<tr>
<td>Q4</td>
<td>1</td>
<td>3.1</td>
<td>8</td>
<td>25.0</td>
<td>11</td>
<td>34.4</td>
<td>12</td>
<td>37.5</td>
<td>32</td>
</tr>
<tr>
<td>Q5</td>
<td>2</td>
<td>6.3</td>
<td>9</td>
<td>28.1</td>
<td>4</td>
<td>12.5</td>
<td>17</td>
<td>53.1</td>
<td>32</td>
</tr>
<tr>
<td>Q6</td>
<td>1</td>
<td>3.1</td>
<td>5</td>
<td>15.6</td>
<td>12</td>
<td>37.5</td>
<td>14</td>
<td>43.8</td>
<td>32</td>
</tr>
<tr>
<td>Q7</td>
<td>1</td>
<td>3.1</td>
<td>9</td>
<td>28.1</td>
<td>13</td>
<td>40.6</td>
<td>9</td>
<td>28.1</td>
<td>32</td>
</tr>
<tr>
<td>Q8</td>
<td>2</td>
<td>6.3</td>
<td>10</td>
<td>31.3</td>
<td>9</td>
<td>28.1</td>
<td>11</td>
<td>34.4</td>
<td>32</td>
</tr>
<tr>
<td>Q9</td>
<td>3</td>
<td>9.4</td>
<td>8</td>
<td>25.0</td>
<td>8</td>
<td>25.0</td>
<td>13</td>
<td>40.6</td>
<td>32</td>
</tr>
<tr>
<td>Q10</td>
<td>1</td>
<td>3.1</td>
<td>12</td>
<td>37.5</td>
<td>10</td>
<td>31.5</td>
<td>9</td>
<td>28.1</td>
<td>32</td>
</tr>
</tbody>
</table>

Note. F – Frequency. % – Percentage.

Table 4. Frequency table for experimental group

<table>
<thead>
<tr>
<th>Likert scale</th>
<th>3</th>
<th>%</th>
<th>4</th>
<th>%</th>
<th>5</th>
<th>%</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>37.5</td>
<td>15</td>
<td>62.5</td>
<td>24</td>
</tr>
<tr>
<td>Q2</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>37.5</td>
<td>15</td>
<td>62.5</td>
<td>24</td>
</tr>
<tr>
<td>Q3</td>
<td>2</td>
<td>8.3</td>
<td>14</td>
<td>58.3</td>
<td>8</td>
<td>33.3</td>
<td>24</td>
</tr>
<tr>
<td>Q4</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>33.3</td>
<td>16</td>
<td>66.7</td>
<td>24</td>
</tr>
<tr>
<td>Q5</td>
<td>2</td>
<td>8.3</td>
<td>11</td>
<td>45.8</td>
<td>11</td>
<td>45.8</td>
<td>24</td>
</tr>
<tr>
<td>Q6</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>37.5</td>
<td>15</td>
<td>62.5</td>
<td>24</td>
</tr>
<tr>
<td>Q7</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>45.8</td>
<td>13</td>
<td>54.2</td>
<td>24</td>
</tr>
<tr>
<td>Q8</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>37.5</td>
<td>15</td>
<td>62.5</td>
<td>24</td>
</tr>
<tr>
<td>Q9</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>37.5</td>
<td>15</td>
<td>62.5</td>
<td>24</td>
</tr>
<tr>
<td>Q10</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>29.2</td>
<td>17</td>
<td>70.8</td>
<td>24</td>
</tr>
</tbody>
</table>

Note. F – Frequency. % – Percentage.

Table 5. Reliability of Cognitive Load instrument based on the Cronbach’s alpha

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Experimental</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cronbach’s alpha</td>
<td>.8604</td>
<td>.7704</td>
</tr>
</tbody>
</table>

In the case of usability, the analysis reflects the mean SUS score (M) for each one of the groups (Table 6). In particular, the mean SUS score is shown for each group, which can be considered acceptable (SUS > 70) for both groups, being in the case of the control group on the grade scale of “good” (SUS > 70) and in the case of the experimental group on the grade scale near “excellent” (SUS > 80). Comparing these usability values, we can see that the usability of the integrated system (LMS endowed with Cloud services) built through our semantic description method is further improved, which allows students to have a more satisfying experience in the educational scenario they participated in terms of usability.

Table 6. Result value of SUS for Control Group and Experimental Group

<table>
<thead>
<tr>
<th></th>
<th>Control Group</th>
<th>Experimental Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>32</td>
<td>24</td>
</tr>
<tr>
<td>max</td>
<td>95</td>
<td>95</td>
</tr>
<tr>
<td>Min</td>
<td>57.5</td>
<td>75</td>
</tr>
<tr>
<td>M</td>
<td>76</td>
<td>83.3</td>
</tr>
<tr>
<td>SD</td>
<td>8.25</td>
<td>8.82</td>
</tr>
</tbody>
</table>

Note. n = students; max = maximum value; min = minimum value; M = Mean; SD = standard deviation.

Findings: Response to the research questions

Based on the above results, we can provide solid responses to the two research questions set in the beginning of this paper. In the case of cognitive load / satisfaction relationship, we can argue that the experimental group has better...
results for the indicators of awareness, challenge and engagement than the control group. For each of the questions related to these indicators, the control group students describe their experience as less satisfactory with respect to the experimental group that used the integrated system. Students’ satisfaction in the experimental group has, in turn, a positive impact on cognitive load and subsequently on their learning. Furthermore, with regard to usability, both groups consider the development of the experiment as an acceptable experience, though the experimental group expresses greater satisfaction. From these results we can conclude that an LMS integrated with Cloud services is more acceptable and useful than one in which users have to use them as separate tools.

Conclusions and future work

In this paper we have shown the benefits that users can obtain from exploiting the integration capabilities between an LMS and Cloud services by means of a semantic model. On the one hand, we showed how an LMS that integrates with Cloud services, built out of a semantic description method, has a direct positive impact on students’ cognitive load by reducing it and thus increasing students’ satisfaction levels. On the other hand, we highlighted how this system shows better usability than one in which tools are used independently. Future work is on the way to extend our approach by integrating more learning tools with functionalities that are available in the Cloud, applied in different learning scenarios, with the ultimate aim to provide a comparative study of all the possible benefits, advantages as well as possible limitations. In this new study we will explore the impact of integration on faculty as well, considering the barriers that faculty experience in terms of the gamut of tools that are available to them. To this end we will include factors like faculty attitudes, perceptions and experiences as well as how LMS and Cloud-computing can help facilitate cost-effective, innovative solutions for Higher Education.

References


Exploring the Profiles and Interplays of Pre-service and In-service Teachers’ Technological Pedagogical Content Knowledge (TPACK) in China

Yan Dong¹, Ching Sing Chai³, Guo-Yuan Sang²*, Joyce Hwee Ling Koh¹ and Chin-Chung Tsai⁴

¹Collaborative & Innovative Center for Educational Technology, co-supported by Beijing Normal University and Central China Normal University, China // ²Beijing Normal University, #19, Xinwai St., Haidian Dist., Beijing, 100875, China // ³Nanyang Technological University, 1 Nanyang Walk, Singapore // ⁴National Taiwan University of Science and Technology, #43, Sec.4, Keelung Rd., Taipei, 106, Taiwan // Dongy@bnu.edu.cn // Chingsing.chai@nie.edu.sg // guoyuansang@gmail.com // Joyce.koh@nie.edu.sg // cctsai@mail.ntust.edu.tw

*Corresponding author

(Submitted November 7, 2013; Revised February 18, 2014; Accepted March 18, 2014)

ABSTRACT

This research surveyed three hundred and ninety pre-service and three hundred and ninety four in-service teachers with regards to the seven factors of technological pedagogical content knowledge, their beliefs about constructivist oriented teaching (CB) and design disposition (DD). Both exploratory and confirmatory factor analyses showed that the survey based on the nine-factor model had high reliability and validity. Significant differences between pre-service and in-service teachers in the TPACK factors and CB and DD were found and the differences reveal that the pre-service teachers are less knowledgeable and confident with regards to all the factors. In order to identify predictors of TPACK, the research further explores the relationships among TPACK factors, CB and DD through structural equation models. The findings reveal that DD consistently predict both pre-service and in-service teachers’ TPACK and this provide support about the importance of design disposition for TPACK advancement. However, CB does not predict the pre-service teachers’ TPACK. In addition, CB is a significantly negative predictor for the in-service teachers’ TPACK. The findings may imply that while the in-service teachers believe strongly in constructivist oriented teaching, they need further professional development in designing instruction to actualize their desired form of education.

Keyword

Technological pedagogical content knowledge (TPACK), Teachers’ beliefs, Design disposition, Teacher education

Introduction

The theoretical framework of technological pedagogical content knowledge (TPACK) has emerged as a widely accepted framework to provide explanation of teachers’ work in integrating Information and Communication Technologies (ICT) into classroom (Mishra & Koehler, 2006; Cox & Graham, 2009). The TPACK framework accounts for the integration of ICT from a knowledge perspective. It identifies three elementary forms of knowledge that must be present in any ICT integrated lesson: technological knowledge (TK), pedagogical knowledge (PK) and content knowledge (CK).

The interactions of these three forms of knowledge were postulated to develop other secondary forms of knowledge that include pedagogical content knowledge (PCK) (Shulman, 1986), technological pedagogical knowledge (TPK), technological content knowledge (TCK) and ultimately the synthesized form of knowledge, which is the TPACK. Shulman (1986) postulated that PCK is the unique form of teacher knowledge that synthesizes the teacher’s PK and CK to help students in bridging the difficulties of mastering the subject matter. It could be in the form content-specific or topic-specific pedagogical approaches (Graham, Borup, & Smith, 2012). In similar vein, TPK is the unique form of PK that is associated with the use of specific kinds of technology. TCK refers to technologically represented content knowledge that is not created for the purpose of teaching. Finally, TPACK refers to the integrated form of knowledge that can be created through different combination of the six forms of knowledge discuss earlier (Cox et al., 2009). This form of knowledge is created through teachers or educational technologists’ design effort when they generate new practices to integrate ICT into classroom teaching (Tsai, Chai, Wong, Hong, & Tan, 2013). Currently, the notion of TPACK is spreading and being adopted by many countries to understand and enhance teachers’ ability to integrate ICT (Chai, Koh & Tsai, 2013a).
Researchers in China have also begun to explore teachers’ ability to integrate ICT using the TPACK framework. However, the rigor of research may need enhancement (see Zhan & Ren, 2011). In particular, factor analyses were not performed to verify the survey’s factor structure. This research aims to explore the Chinese teachers’ TPACK profiles and the relationships among TPACK factors, teachers’ constructivist beliefs and design disposition through structural equation models.

**Theoretical background**

Given the growing interest in TPACK, a number of studies that aimed to develop a valid and reliable instrument has emerged (Archambault & Crippen; 2009; Chai, Ng, Lee, Hong, & Koh, 2013b). Methodologically, it seems that Chai and his colleagues’ survey instrument possesses good construct validity as revealed through confirmatory and the instrument possesses strong reliabilities. In addition, it fits that Asian context best (see Chai et al., 2013b) and it is based on current emphasis towards constructivist-oriented use of technology in the classrooms.

Theoretically, the elementary forms of knowledge (CK, PK, and TK) and the secondary forms of knowledge (TPK, PCK, and TCK) should act as epistemic resources to support the teachers’ development of TPACK. Past research on structural equation models had provided general support that the elementary forms of knowledge are predictors of teachers’ TPACK (see Chai et al., 2013b; Kramarski, Koh, Tsai, & Tan, 2011). TPACK is thus a situated and complex form of knowledge creation (Mishra & Koehler, 2006; Tee & Lee, 2011) in that the teacher has to consider the students’ readiness, prior knowledge, school technological environments, appropriate teaching methods for the topic to be taught and taught technologies with their specific educational affordances; before the teacher can synthesize these elements into a coherent whole for a specific lesson. However, depending on the design experiences, the level of teacher’s knowledge and the professional development that teachers received, the elementary or secondary forms of knowledge may or may not contribute to the final TPACK. Further reviews about how the elementary factors of knowledge contribute to higher forms of knowledge are discussed in the next paragraphs.

To date, there are only a couple of studies that could map out the relationships between the seven factors through structural equation model (SEM) (see Koh, Chai & Tsai, 2013; Chai et al., 2013b). The relationships are that each of the elementary factors of TPACK constructs may predict the higher forms of TPACK constructs as proposed by Mishra and Koehler (2006). Koh et al. (2013) reported that among the Singaporean in-service teachers, CK and PCK did not positively predict the teachers’ TPACK while the rest of the factors positively predict TPACK. Chai et al.’s (2013b) investigation on Asian pre-service teachers (including Singapore, Hong Kong, Taiwan and Mainland China), however, shows that the direct positive predictors of pre-service teachers’ TPACK were TCK, PCK and TPK, with the elementary TK, CK, and PK having only indirect effects. In sum, the SEMs between the pre-service and in-service teachers may not be the same, which implies that teacher educators need to understand the different efficacies of these two groups of teachers to specifically address their TPACK development needs.

Current studies that aimed at understand teachers’ TPACK and its elementary form of knowledge through quantitative survey either focused on pre-service teachers or in-service teachers. To understand how teachers’ TPACK differ from pre-service to in-service stages necessitate some forms of tracking using similar instrument. Greenhow, Dexter and Hughes (2008) attempted to understand the differences between in-service (n = 17) and pre-service (n = 16) science teachers’ performances and instructional decisions for ICT integration base on multimedia case scenarios. More studies are desirable to establish a more generalizable pattern of teachers’ TPACK development to compare their difference.

In addition, the adequacy of the TPACK framework has been challenged recently by Voogt and Roblin (2012) and Chai et al. (2013a) in their review of TPACK research. These researchers pointed out that teachers’ work has always been shaped by many different strands of beliefs that they hold. The TPACK framework has been thus far focused exclusively on accounting for the knowledge that teachers need to integrate technology. It is therefore necessary to further unpack the relationship between teachers’ TPACK and associated beliefs. There are currently few studies that explicitly examine both teachers’ TPACK and beliefs concurrently with exceptions such as a correlation study conducted by Chai, Chin, Koh, and Tan (2013c).

Among the related beliefs, we focus on teachers’ constructivist beliefs since the integration of ICT is often targeted on students’ center knowledge construction (Jonassen & Ionas, 2008). Constructivist beliefs are also known as
“progressive beliefs” or “student-centred approaches” (Bramald et al., 1995) and are often regarded as beliefs that “support student learning” (Samuelowicz & Bain, 1992) and provide a “constructivist philosophy of learning” (Bramald et al., 1995). Those teachers who focus on constructive and progressive teaching and learning processes adopt constructivist beliefs (Sang, Valeke, van Braak, & Tondeur, 2009). As Tondeur and his colleagues (2008) argue, teachers who are more inclined towards adopting constructivist beliefs are more active ICT users compared with teachers who are less inclined towards adopting the constructivist beliefs.

Furthermore, as TPACK is in essence designed knowledge (Chai et al., 2013a), we argue that teachers’ disposition towards design is an important aspect of teachers’ beliefs that need attention. Design disposition is constituted through the expectations and inclinations of the designer that could influence the design activity (Boland & Collopy, 2004). Expert designers have been found to possess distinctive dimensions of design disposition. These dimensions are the capacity to embrace ambiguity and openness; develop emphatic understanding of user’s needs; attend to multidimensional meanings; generating polysensorial aesthetics and creating artefacts and solutions to address the design problems (Michlewski, 2008). Cross (2007) highlighted tolerance towards uncertainty and conflicting ideas and patience towards the emergence of solutions as essential dispositions of designers. In sum, researchers who have studied designers’ disposition have identified some intra-personal psychological traits needed for good designers. Traditionally, there may be a higher proportion of teachers who practice teaching as delivery of information rather than designing the learning environment and activities to facilitate students’ knowledge construction. However, rapid technological changes challenge the stasis in curriculum and instructional design. In fact, the lack of design capacity has been dubbed as the third order barrier to ICT integration (Tsai & Chai, 2012); following after the first order barrier of access to technology and the second order of teachers’ beliefs. We thus propose in this study to investigate the hypothesis that teachers’ with stronger design disposition would be more inclined to develop TPACK.

Based on the foregoing review, we hypothesize that the elementary and secondary forms of knowledge would predict TPACK and that the teachers' constructivist beliefs and design disposition would positively predict their TPACK. Figure 1 below depicts the theoretical model formed. The hypotheses (H) tested in this model include:

H1 CK predicts TCK; H2 CK predicts PCK; H3 CK predicts TPACK; H4 TK predicts TPK; H5 TK predicts TCK; H6 TK predicts TPACK; H7 PK predicts TPK; H8 PK predicts PCK; H9 PK predicts TPACK; H10 TCK predicts TPACK; H11 TPK predicts TPACK; H12 PCK predicts TPACK; H13 CB predicts TPACK; H14 DD predicts TPACK.

![Figure 1. Theoretical model proposed in this paper](image)

**Research questions**

Based on the reviewed literature, our study aims to address the following research questions:
Do the nine-factor model that includes the seven TPACK constructs, teachers’ constructivist beliefs (CB) and design disposition (DD) fit the Chinese teacher context? Are there any differences between pre-service and in-service teachers’ TPACK and teachers associated CB and DD? Do Chinese pre-service and in-service teachers’ TPACK primitive constructs (i.e., TK, PK, CK, TCK, TPK, PCK), CB and DD predict their TPACK?

For research question 1, it was hypothesized that the contextualization of the TPACK survey for Chinese pre-service and in-service teachers would yield satisfactory fit indices for the combined model with the seven TPACK factors as postulated by Mishra & Koehler (2006) and CB and DD proposed in this research. For research question 2, it was hypothesized that Chinese teachers had different TPACK profiles with other contexts and that there were some differences between pre-service and in-service teachers’ TPACK. For research question 3, it was hypothesized that CB and DD had both direct contributions to the TPACK as illustrated in Figure 1.

**Method**

**Participants**

The participants for our study were 390 pre-service (M age = 20.59, SD = 1.661) and 394 in-service teachers (M age = 36.06, SD = 7.226) in Beijing (China). The sample of pre-service teachers in this survey are undergraduates from 11 provinces of China and they will have to go back to their respective province to be primary or secondary school teachers after graduation, according to related policies. Pre-service teachers are majoring six subjects: Chinese Language, English Language, History, Education, Math and Physics.

| Table 1. Sample distribution of pre-service and in-service teachers (N = 784) |
|---------------------------------|---|------------------|---|------------------|
|                               | Options | Number (%) | Options | Number (%) |
| Pre-service                    |         |            | In-service |         |
| All sample                     | 390 (100) | 394 (100) |
| Gender                         |         |            | Gender     |         |
| Male                           | 131 (33.6) | 81 (20.6) |
| Female                         | 259 (66.4) | 310 (78.7) |
| Grade                          |         |            | Teaching   |         |
| Senior                         | 166 (42.6) | 57 (14.5) |
| Junior                         | 77 (19.7) | 160 (40.6) |
| Sophomore                      | 89 (22.8) | 173 (43.9) |
| Freshmen                       | 58 (14.9) |            | > 16 years |         |
| Family                         |         |            | < 5 years  |         |
| Rural                          | 206 (52.8) | 243 (61.7) |
| Location                       |         |            | 6-15 years |         |
| Urban                          | 172 (44.1) | 119 (30.2) |

For pre-service teachers (n = 390), 259 (66.4%) were female. Most of the respondents were in their last year of college (166, 42.6%). Freshmen were the smallest population of the respondents (58, 14.9%). As to family locations of pre-service teachers, most of them came from rural areas (206, 52.8%), and 172 (44.1%) came from urban areas.

Of all respondents in in-service teachers (n = 394), 78.7% teachers were female. Respondents were grouped into 3 categories depending their years of teaching experience: teachers with less than 5 years of teaching experience (14.5%); teachers with 6-15 years of teaching experience (40.6%); and teachers with more than 16 years of teaching experience (43.9%). Furthermore, schools of 243 (61.7%) teachers were located in rural areas, and 119 (30.2%) were located in urban areas.

**Instrument**

The instrument was adapted from several sources. The TPACK constructs and Constructivism Beliefs were adapted from Chai et al. (2013b) and Chai et al. (2013c) respectively. The 6 items on design deposition were taken from Koh, Chai, Hong and Tsai (2013). A total of 46 items were assembled and they were subsequently translated by two professors majoring in Chinese language. Therefore, the final survey comprised of 46 items that measured teachers’ perceptions with respect to the extent to which they agreed or disagreed with statements related to the seven TPACK constructs and their beliefs with respect to the constructivism and design deposition. Each statement was rated as a
seven point Likert-type scale with 1-strongly disagree, 2-disagree, 3-slightly disagree, 4-neutral, 5-slightly agree, 6-agree, 7-strongly agree.

Data collection and analysis

Paper-based surveys were distributed to the participants in the schools for the in-service teachers and during class for the pre-service teachers. The surveys were administered anonymously. The teachers took around 15-20 minutes to complete the surveys. The responses were then collated in SPSS. Since only the TPACK portion survey has been previously validated by Asian pre-service teachers which include China sample, confirmatory factor analysis was performed to ensure that the 9 factors model is valid using AMOS 21. Eight items were deleted including TPK1, TPK2, TPACK2, PK1, PCK1, PCK2, TCK2, DD2 because these items showed cross loadings or low loadings. Internal reliability was first established through high overall Cronbach’s alpha for the survey is 0.98 and the respective Cronbach’s alphas are: TK (α = 0.95), PK (α = 0.94), CK (α = 0.89), PCK (α = 0.95), TPK (α = 0.91), TCK (α = 0.88), TPACK (α = 0.91), CB (α = 0.95), DD (α = 0.89). Further analyses based on the model include bivariate correlations, t-tests and testing of the structural equation model based on the identified factors.

Results

Measurement model and independent t-tests

The CFA performed on the measurement model includes TPACK seven constructs and CB and DD. The CFA obtained satisfactory model fit ($\chi^2 = 2040.63$, $\chi^2$/df = 2.91, $p < 0.001$, TLI = .95, CFI = .96, RMSEA = 0.049 (LO90 = 0.047, HI90 = 0.052), SRMR = 0.071). Table 2 shows that nine constructs’ mean score, SD, t-value between pre-service teachers and in-service teachers. Table 2 also shows that there are significant differences ($p < .001$) between pre-service and in-service teachers in the TPACK constructs and CB and DD. TPACK framework constructs and associated beliefs (CB and DD) of pre-service teachers are lower than those of in-service teachers, especially in the elementary factors (TK, CK, PK). The results also indicate that CK, regardless of pre-service teachers or teachers, is the lowest among CK, TK and PK. CB is higher than others, implying that pre-service and in-service teachers have relatively strong constructivist beliefs.

<table>
<thead>
<tr>
<th>Type</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
<th>t-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TK</td>
<td>Pre-S-T</td>
<td>390</td>
<td>4.48</td>
<td>1.24</td>
<td>.063</td>
</tr>
<tr>
<td></td>
<td>In-S-T</td>
<td>394</td>
<td>5.44</td>
<td>1.20</td>
<td>.060</td>
</tr>
<tr>
<td>CK</td>
<td>Pre-S-T</td>
<td>390</td>
<td>4.01</td>
<td>1.20</td>
<td>.061</td>
</tr>
<tr>
<td></td>
<td>In-S-T</td>
<td>394</td>
<td>5.19</td>
<td>1.16</td>
<td>.058</td>
</tr>
<tr>
<td>PK</td>
<td>Pre-S-T</td>
<td>390</td>
<td>4.60</td>
<td>1.09</td>
<td>.056</td>
</tr>
<tr>
<td></td>
<td>In-S-T</td>
<td>394</td>
<td>5.63</td>
<td>0.97</td>
<td>.049</td>
</tr>
<tr>
<td>PCK</td>
<td>Pre-S-T</td>
<td>390</td>
<td>4.57</td>
<td>1.10</td>
<td>.056</td>
</tr>
<tr>
<td></td>
<td>In-S-T</td>
<td>394</td>
<td>5.20</td>
<td>1.18</td>
<td>.059</td>
</tr>
<tr>
<td>TPK</td>
<td>Pre-S-T</td>
<td>390</td>
<td>4.71</td>
<td>1.10</td>
<td>.055</td>
</tr>
<tr>
<td></td>
<td>In-S-T</td>
<td>394</td>
<td>5.23</td>
<td>1.30</td>
<td>.065</td>
</tr>
<tr>
<td>TCK</td>
<td>Pre-S-T</td>
<td>390</td>
<td>4.43</td>
<td>1.09</td>
<td>.055</td>
</tr>
<tr>
<td></td>
<td>In-S-T</td>
<td>394</td>
<td>5.24</td>
<td>1.26</td>
<td>.063</td>
</tr>
<tr>
<td>TPACK</td>
<td>Pre-S-T</td>
<td>390</td>
<td>4.41</td>
<td>1.07</td>
<td>.054</td>
</tr>
<tr>
<td></td>
<td>In-S-T</td>
<td>394</td>
<td>5.03</td>
<td>1.28</td>
<td>.064</td>
</tr>
<tr>
<td>CB</td>
<td>Pre-S-T</td>
<td>390</td>
<td>5.50</td>
<td>1.25</td>
<td>.063</td>
</tr>
<tr>
<td></td>
<td>In-S-T</td>
<td>394</td>
<td>6.11</td>
<td>1.05</td>
<td>.053</td>
</tr>
<tr>
<td>DD</td>
<td>Pre-S-T</td>
<td>390</td>
<td>4.61</td>
<td>0.96</td>
<td>.048</td>
</tr>
<tr>
<td></td>
<td>In-S-T</td>
<td>394</td>
<td>5.30</td>
<td>1.02</td>
<td>.051</td>
</tr>
</tbody>
</table>

Note. *** $p < 0.001$.

Correlation analysis
Significant positive correlations between the TPACK constructs were established at $p < 0.01$ (See Table 3), indicating that these relationships could be further examined with SEM. TCK and TPK had the largest positive correlations with TPACK while PK had relatively lower but positive significant correlation with TPACK. The correlation between CB and TPACK was lower than that between DD and TPACK, which was around 0.60. Similarly, the correlations between CB and other constructs were lower than the correlations between DD and those constructs.

Table 3. Inter-correlations between TPACK constructs and CB, DD

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[1]CK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-service</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In-service</td>
<td>.69**</td>
<td>.59**</td>
<td>.55**</td>
<td>.58**</td>
<td>.51**</td>
<td>.50**</td>
<td>.51**</td>
<td>.47**</td>
</tr>
<tr>
<td>[2]PK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-service</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In-service</td>
<td>.59**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[3]TK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-service</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In-service</td>
<td>.55**</td>
<td>.58**</td>
<td>.46**</td>
<td>.49**</td>
<td>.49**</td>
<td>.62**</td>
<td>.51**</td>
<td>.45**</td>
</tr>
<tr>
<td>[4]PCK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-service</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In-service</td>
<td>.58**</td>
<td>.57**</td>
<td>.46**</td>
<td>.49**</td>
<td>.55**</td>
<td>.48**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[5]TPK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-service</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In-service</td>
<td>.51**</td>
<td>.55**</td>
<td>.69**</td>
<td>.50**</td>
<td>.68**</td>
<td>.57**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[6]TCK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-service</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In-service</td>
<td>.50**</td>
<td>.51**</td>
<td>.72**</td>
<td>.45**</td>
<td>.68**</td>
<td>.52**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[7]TPACK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-service</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In-service</td>
<td>.51**</td>
<td>.56**</td>
<td>.71**</td>
<td>.52**</td>
<td>.77**</td>
<td>.52**</td>
<td>.71**</td>
<td>.56**</td>
</tr>
<tr>
<td>[8]CB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-service</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In-service</td>
<td>.47**</td>
<td>.61**</td>
<td>.51**</td>
<td>.44**</td>
<td>.40**</td>
<td>.50**</td>
<td>.40**</td>
<td>.39**</td>
</tr>
<tr>
<td>[9]DD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-service</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In-service</td>
<td>.15**</td>
<td>.42**</td>
<td>.34**</td>
<td>.43**</td>
<td>.56**</td>
<td>.39**</td>
<td>.56**</td>
<td>.39**</td>
</tr>
</tbody>
</table>

Note. $N_{\text{Pre-service}} = 390$, $N_{\text{In-service}} = 394$. ** $p < 0.01$.

Structural equation model

The results of the structural model based on the hypotheses (displayed in Figure 1) are presented in Figure 2 for pre-service teachers and Figure 3 for in-service teachers. The SEM yielded satisfactory model fit shown in Table 4. The path coefficients are summarized in Table 5 for pre-service and Table 6 for in-service teachers. In the figures, straight line indicates a significant relationship studied in the literature; dashed line indicates a suggested but empirically insignificant relationship in this study.

Figure 2. CB, DD and Pre-service Teachers’ TPACK
From Figure 2 and Figure 3, it can be seen that, by and large, most pathways to TPACK as proposed were supported.

**Table 4.** Model fit summary of pre-service and in-service teachers

<table>
<thead>
<tr>
<th>Model</th>
<th>χ²</th>
<th>Df</th>
<th>χ²/df</th>
<th>IFI</th>
<th>TLI</th>
<th>CFI</th>
<th>RMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-service</td>
<td>1586.00</td>
<td>712</td>
<td>2.23</td>
<td>.93</td>
<td>.92</td>
<td>.93</td>
<td>.056</td>
</tr>
<tr>
<td>Teachers</td>
<td>In-service</td>
<td>1845.99</td>
<td>712</td>
<td>2.59</td>
<td>.93</td>
<td>.92</td>
<td>.93</td>
</tr>
</tbody>
</table>

**Table 5.** Path coefficients of the TPACK SEM (for pre-service teachers)

<table>
<thead>
<tr>
<th>Hypotheses</th>
<th>Path</th>
<th>Estimate</th>
<th>SE</th>
<th>C.R.</th>
<th>p values</th>
<th>Supported ?</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>CK → TCK</td>
<td>.13</td>
<td>.04</td>
<td>3.28</td>
<td>**</td>
<td>Yes</td>
</tr>
<tr>
<td>H2</td>
<td>CK → PCK</td>
<td>.09</td>
<td>.05</td>
<td>1.70</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>H7</td>
<td>PK → TPK</td>
<td>.35</td>
<td>.05</td>
<td>6.36</td>
<td>***</td>
<td>Yes</td>
</tr>
<tr>
<td>H8</td>
<td>PK → PCK</td>
<td>.64</td>
<td>.07</td>
<td>8.65</td>
<td>***</td>
<td>Yes</td>
</tr>
<tr>
<td>H5</td>
<td>TK → TPK</td>
<td>.46</td>
<td>.05</td>
<td>9.98</td>
<td>***</td>
<td>Yes</td>
</tr>
<tr>
<td>H4</td>
<td>TK → TCK</td>
<td>.63</td>
<td>.05</td>
<td>12.55</td>
<td>***</td>
<td>Yes</td>
</tr>
<tr>
<td>H3</td>
<td>CK → TPACK</td>
<td>-.12</td>
<td>.07</td>
<td>-1.80</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>H9</td>
<td>PK → TPACK</td>
<td>.05</td>
<td>.07</td>
<td>.72</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>H6</td>
<td>TK → TPACK</td>
<td>-.00</td>
<td>.07</td>
<td>-.02</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>H11</td>
<td>TPK → TPACK</td>
<td>.31</td>
<td>.06</td>
<td>5.26</td>
<td>***</td>
<td>Yes</td>
</tr>
<tr>
<td>H10</td>
<td>TCK → TPACK</td>
<td>.49</td>
<td>.07</td>
<td>7.05</td>
<td>***</td>
<td>Yes</td>
</tr>
<tr>
<td>H12</td>
<td>PCK → TPACK</td>
<td>-.05</td>
<td>.04</td>
<td>-1.21</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>H13</td>
<td>CB → TPACK</td>
<td>-.01</td>
<td>.03</td>
<td>-.25</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>H14</td>
<td>DD → TPACK</td>
<td>.12</td>
<td>.06</td>
<td>2.14</td>
<td></td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Note.*** p < 0.001. ** p < 0.01. * p < 0.05.

**Table 6.** Path coefficients of the TPACK SEM (for in-service teachers)

<table>
<thead>
<tr>
<th>Hypotheses</th>
<th>Path</th>
<th>Estimate</th>
<th>SE</th>
<th>C.R.</th>
<th>p values</th>
<th>Supported ?</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>CK → TCK</td>
<td>.18</td>
<td>.06</td>
<td>2.47</td>
<td>**</td>
<td>Yes</td>
</tr>
<tr>
<td>H2</td>
<td>CK → PCK</td>
<td>.47</td>
<td>.10</td>
<td>4.86</td>
<td>***</td>
<td>Yes</td>
</tr>
<tr>
<td>H7</td>
<td>PK → TPK</td>
<td>.24</td>
<td>.07</td>
<td>3.49</td>
<td>***</td>
<td>Yes</td>
</tr>
<tr>
<td>H8</td>
<td>PK → PCK</td>
<td>.26</td>
<td>.11</td>
<td>2.35</td>
<td>*</td>
<td>Yes</td>
</tr>
<tr>
<td>H5</td>
<td>TK → TPK</td>
<td>.66</td>
<td>.05</td>
<td>12.63</td>
<td>***</td>
<td>Yes</td>
</tr>
<tr>
<td>H4</td>
<td>TK → TCK</td>
<td>.72</td>
<td>.06</td>
<td>12.78</td>
<td>***</td>
<td>Yes</td>
</tr>
<tr>
<td>H3</td>
<td>CK → TPACK</td>
<td>-.12</td>
<td>.07</td>
<td>-1.80</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>H9</td>
<td>PK → TPACK</td>
<td>.07</td>
<td>.08</td>
<td>.87</td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>H6</td>
<td>TK → TPACK</td>
<td>-.03</td>
<td>.07</td>
<td>-.42</td>
<td></td>
<td>No</td>
</tr>
</tbody>
</table>
Figure 2 depicts the path coefficients that significantly predict TPACK constructs of pre-service teachers. The results indicated that teachers perceived three direct paths to TPACK from CK, TCK and TPK. They also perceived DD as predictor of their TPACK (See Figure 2). However, TCK and TPK had larger path coefficients, indicating that these were perceived to have stronger prediction on TPACK than CK and DD. Among these four predictors, TCK had the largest path coefficient. The results indicated that TK and PK were perceived to predict teachers’ PCK. However, no significant path between PCK and TPACK was established, indicating that the teachers did not perceive significant relationships between their PCK and TPACK. The direct pathways of CK on PCK and CB on TPACK were also not significant.

Figure 3 depicts the path coefficients of significant paths towards TPACK constructs of in-service teachers. The study results indicated that teachers perceived four direct paths to TPACK from TCK and TPK. They perceived both CB and DD to be predictors of their TPACK (See Figure 3). However, TCK and TPK had larger path coefficients, indicating that these were perceived to be stronger predictors of TPACK than CB and DD. Among these four direct paths, TCK had the largest path coefficient. The study results indicated that TK and PK were perceived to be predictors of teachers’ PCK. However, no significant path between PCK and TPACK was established, indicating that the in-service teachers did not perceive significant relationships between their PCK and TPACK.

Discussion

This study indicates that the nine factors survey we adapted to measure the Chinese pre-service and in-service teachers’ TPACK and associated beliefs possesses construct validity as assessed through CFA. The reliabilities were also strong. This finding advances research of TPACK in that it has provided a possible instrument to measure teachers’ beliefs (CB and DD) and their TPACK concurrently. The validation of the 9 factors TPACK and CB and DD survey helps researchers in unpacking the complex relationships among these constructs. This research therefore is based on Chai et al.’s (2011, 2013b, 2013c) research by an expanded model of teachers’ TPACK that considers not only the 7 factors of knowledge but it also included important beliefs associated with TPACK. In addition, the additional beliefs were examined through SEM to investigate whether or not they are significant predictors (see later). In particular, Chai et al. (2013c) began the work of investigating teachers’ beliefs with TPACK through correlations and it is conducted specifically for Singaporean Chinese language teacher. It is therefore a subject specific study. This study involves a larger sample of both in-service and pre-service teachers from China and the teachers specialized in different subject matter teaching.

Table 2 provides the mean scores and standard deviations for each factor measured with two groups. Pre-service teachers’ TPACK constructs and associated beliefs were all lower than in-service teachers. In terms of the pre-service teachers’ TPACK profile, it shows that the teachers perceived themselves strongest in term of their TPK (M = 4.71, SD = 1.09) and weakest in terms of their CK (M = 4.01, SD = 1.20). All other factors of TPACK fall within these range, indicating that the pre-service teachers do not possess high efficacy about TPACK knowledge. It seems necessary for educational institutions in China to review the related technology courses and engage pre-service teachers to integrate ICT through design. Adaptations of reported pre-service teacher curriculum targeting at enhancing pre-service teachers ICT integration could be helpful (see Angeli & Valanides, 2013; Chai et al., 2011; Zhan et al., 2011). The method of technology mapping could improve the pre-service teachers’ design capacities.

In terms of the in-service teachers’ TPACK profile, it shows that the teachers perceived themselves strongest in term of their PK (M = 5.63, SD = .97) and weakest in terms of their TPACK (M = 5.03, SD = 1.28). The profile we obtained through the survey indicated that Chinese in-service teachers are at least with some confidence in their TPACK but it also reveals that there is a clear need to further develop the teachers’ TPACK, especially through professional development activities that directly engage them in designing ICT integrated lesson (Mishra & Koehler, 2006).
Engaging in-service teachers to work in groups to design instruction collaboratively could provide the necessary means to improve their TPACK (see Koehler, Mishra & Yahya, 2007).

With regards to the beliefs measured, both the pre-service and the in-service teachers are inclined toward constructivist beliefs, with the in-service teachers expressing stronger inclinations toward CB. These findings are similar to Sang and his colleagues’ studies in Chinese educational settings (e.g., Sang et al., 2009; Sang, Valcke, Tondeur, van Braak, & Zhu, 2012). As for their DD, both groups does not seem to be adverse about design problems but their comfort level may need to be enhanced as integrating ICT into classroom is in essence dealing with the wicked problems associated with design (Chai et al., 2011). It may be beneficial to explicitly point out to the teachers that the epistemic nature of creating ICT integration lesson is knowledge creation in a designedly ways of knowing (Chai et al., 2011; Tee et al., 2011); and such endeavor involves dealing with multiple possible solutions without clear criteria about which solution is superior. The assessment about the quality of the designed solution is a situated form of knowledge critique that has to be socially constructed. When one is conscious that one is engaged in knowledge creation work, one may better appreciate the uncertainty involves and adjust one's expectation accordingly. Further research on how teacher’s consciousness about the epistemic nature of creating TPACK influence their design cognition; in comparison with those teachers who are not conscious about it; would deepen our understanding about TPACK creation.

The hypotheses were depicted in Figure 1 and we predicted based on previous theory and research that the elementary forms of knowledge are knowledge resources that contribute to the teachers TPACK and that CB and DD would contribute teachers’ TPACK. We discussed the hypotheses concerning the TPACK factors first before moving into CB and DD. In general, the SEMs obtained support that the elementary forms (CK, PK, TK) predict the secondary forms of knowledge (TPK, TCK, PCK), and TCK and TPK predict TPACK. However, PKC did not predict TPK for both groups of teachers. For the pre-service teachers, CK did not predict PCK, and TK and PK also did not predict TPACK directly. In addition, PKC did not predict TPACK, but CK directly predict TPACK. The structural equation models that we have obtained bear similarity and differences as compared to what has been published to date. Chai et al. (2013b) obtained similar results among Asian pre-service teachers, which include another sample of Chinese pre-service teachers from another region. However, PCK was a significant predictor of TPACK for Chai et al.’s (2013b) study.

For the in-service teachers, CK predicts PCK and TCK, but did not directly predict TPACK. Similar findings were obtained when it is compared to Singapore in-service teachers except that for Singapore in-service teachers (see Koh et al., 2013). Considering these three studies, it can be seen that elementary TPACK factors (CK, PK, TK) may or may not predict the final factors directly; whereas TPK and TCK consistently predict TPACK (see Koh, Chai & Tsai, 2014). Whether or not PCK positively predict TPACK is dependent on the specific samples. In any case, it seems that teachers did not perceive PCK as a strong positive predictor of TPACK when the path coefficients are taken into account. It seems that more work on helping teachers to link their established PCK with the technology related factors is important. The finding we obtained could signify that once PCK is formed, teachers tend not make conscious effort in building the technology to it.
When the two models are compared, it seems that the in-service teachers have a model that more closely reflects the theoretical model while the pre-service model seems more fragmented. This is a reasonable outcome as the in-service teachers are likely to have more experience in synthesizing TPACK or PCK since designing or adapting lessons is part of teachers’ work. This study contributes to TPACK literature in providing the SEM models for tangible comparison to be made.

Finally, this study identifies the CB and DD as other factors that could predict teachers TPACK. The findings we obtained indicate that DD is a positive predictor for both teachers, but CB is only a negative predictor for in-service teachers. The results are consistent with our research hypothesis that teachers’ DD can predict their TPACK’s development, but it did not support the hypothesis of CB predicting TPACK. The case for DD provides some support to the argument that teacher’s design capacity could be a barrier to ICT integration (Tsai & Chai, 2012). Furthermore, the findings provide support for teacher educators to consider ways of enhancing the teachers’ design position. Explicit discussion about how teachers should handle classroom situations that call for design and scaffolding teachers with general heuristics or cognitive prompts could be helpful (see Kramarski & Michalsky, 2010).

For the unexpected finding of CB being a significant negative predictor of TPACK for the China in-service teachers from Beijing, further discussion is needed. The findings we obtained indicate that the teachers possess high level of constructivist beliefs but the beliefs are significant negative predictors of the teachers’ TPACK. This may indicate that while the teachers believe strongly in constructivist teaching, they are conscious that they lack the design capacity to craft constructivist oriented TPACK. The TPACK items we used in this research refer to teachers’ ability to use technology for student-centered learning for specific subject matters. Sample items include “I can formulate in-depth discussion topics about the content knowledge and facilitate students’ online collaboration with appropriate tools.” Pedagogical activities such as these are student-centered and are usually undergirded by constructivist beliefs towards education (Howland, Jonassen, & Marra, 2012). Past research has also documented that teachers with constructivist beliefs are likely to advocate constructivist oriented ICT-based lesson (Chai, 2010; Sang, Valcke, van Braak, & Tondeur, 2010). However, whether or not they perceived themselves as capable in designing such research was less researched. The finding indicates that possessing constructivist beliefs may not be a sufficient condition for teachers’ TPACK development. Moreover, if adequate professional development is not provided, the in-service teachers’ CB may become a barrier for them to implement some lesson plans that require TPACK. This research has therefore contributed to current literature about teachers’ beliefs and their perceived competency. While beliefs and knowledge are usually correlated, they are not the same thing. Beliefs denote more of an ideal situation while knowledge is akin to actual situation. Our findings may indicate a situation where the teachers wish to conduct ICT integrated lesson supported by technology but they are not sure that they are able to design such lesson. Again, there is a clear need for further development of teachers’ design capacity.

Overall, the identification of two additional predictors for TPACK imply that developing teachers’ TPACK from the perspective of equipping the teachers with relevant knowledge and design experience could be further enhance by attention towards beliefs about themselves as designer (DD) and their teaching beliefs (CB) (see Voogt et al., 2013).

**Conclusion, implications, and limitations**

This study confirms that the nine-factor model that includes the seven TPACK constructs, teachers’ constructivist beliefs and design disposition can be used to explore the Chinese teachers’ TPACK profiles and their development. The 9-factor-model instrument has been developed and validated for both the pre-service and in-service teachers. This study indicated that there were significant differences between these teachers’ TPACK and that design disposition had positive contribution to the TPACK for all teachers, but constructivist beliefs was only a negative predictor for in-service teachers.

There are several implications for teacher education in China. A better understanding of teacher’s beliefs and TPACK can help to improve the efficiency of teacher education programs. An important goal of teacher education programs should consequently be to help teachers developing beliefs that are consistent with the needs of the current or new educational system related to educational technology. In addition, considering the differences between pre-service and in-service teachers, teacher preparation courses should pay much attention to improve student teachers’ knowledge and skill and educational technology and ICT integration.
This study results should be considered in light of several methodological limitations. Firstly, considering difficulties of understanding beliefs and knowledge, quantitative method is not the only way of exploring TPACK and teacher beliefs. More research through qualitative approach is needed to further unpack the above differences and to provide rich data for the refined understanding of the situation. The qualitative case study approach proposed by Jimoyiannis (2010) may be adapted for this purpose. Secondly, although teachers’ design disposition predicts their TPACK, intervention study is needed to discover causal relationship between them. Thirdly, the results and the model proposed fit only in the Chinese teachers; comparative studies are needed to compare with counterparts in different educational contexts.

Acknowledgements

Funding of this research work was supported by the National Educational Science Planning Council, China, under grant number ECA080289.

References


Blending Face-to-Face Higher Education with Web-Based Lectures: Comparing Different Didactical Application Scenarios

Hannelore Montrieux¹*, Sandra Vangestel², Annelies Raes¹, Paul Matthys², Tammy Schellens¹

¹Department of Educational Studies, Faculty of Psychology and Educational Sciences, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium // ²Logistic Department for Teaching, Faculty of Sciences, Ghent University, K.L. Ledeganckstraat 35, 9000 Ghent, Belgium // Hannelore.Montrieux@UGent.be // Sandra.Vangestel@UGent.be // Annelies.Raes@UGent.be // Paul.Matthys@UGent.be // Tammy.Schellens@UGent.be

*Corresponding author

(Submitted October 24, 2013; Revised April 30, 2014; Accepted June 20, 2014)

ABSTRACT
Blended learning as an instructional approach is getting more attention in the educational landscape and has been researched thoroughly. Yet, this study reports the results of an innovation project aiming to gain insight into three different scenarios of applying web-based lectures: as preparation for face-to-face practical exercises, as a repetition of parts of the course, and as an extension of the course. Both the benefits regarding students’ satisfaction, perceived educational effects and the effect on student characteristics were investigated. Results indicate that all students evaluate web-based technology as an added value in higher education. Yet, comparison of the three application scenarios showed that web-based lectures are evaluated most positively when used as course preparation whereas when used as course extension, they were appreciated the least. However, there are indications that student characteristics could play a significant role in the level of appreciation of web-based lectures. We conclude that the potential of web-based lectures lies in their adaptability for use in several scenarios according to educational goals.

Keywords
Educational technology, Blended learning, Web-based lectures, Higher education, Student characteristics

Introduction
In the 21st century, a combination of increasing numbers and increasing diversity of the student body in higher education is a worldwide phenomenon (Preston et al., 2010). This is also the case for Flanders, where a report of the Flemish Ministry of Education found not only an increase in higher education intakes but also an increased diversity of students with regard to background, learning style and learning needs (Flemish Ministry of Education, 2012; Verliefde, Vermeyen, & Van Den Bossche, 2010). Academic staff often feels the need for support in teaching these larger and more diverse groups of students (Office for educational quality control at Ghent University, 2012). In this context, there is an interest in increasing flexibility in higher education, which in 2004 led to a Flemish government decree on the flexibility of the organization of higher education. As a consequence of this demand for flexibility, many universities have introduced web-based lectures, comprising learning tools which integrate sound and images and are designed to digitally record lectures for subsequent delivery over the web (Collis & Moonen, 2011; Gosper, Green, McNeil, Phillips, Preston, & Woo, 2008, von Konisky, Ivins, & Gribble, 2009).

Results of previous academic research demonstrates that web-based lectures are most beneficial regarding student’ satisfaction and learning outcomes when implemented from a blended learning approach, compared with both traditional face-to-face lectures and fully online modes of education (Day, 2008; Day and Foley, 2006; Howlett et al., 2011; Lim & Morris, 2009; Owston, York, & Murtha, 2013; Taradi, Taradi, Radic, & Pokrajac, 2005). Blended learning is an instructional approach that relies on the mixture of face-to-face and web-based learning environments (Garrison & Kanuka, 2004; Graham; 2006). More specific, research by Chen & Liu (2008) found that dynamic media presentations increase learning efficiency and research of Lai, Tsai, & Yu (2011), in which synchronization of a teacher’s lecturing actions for a PowerPoint presentation with his/her voice creates web-based multimedia material which students can use to access past lectures, revealed that students using this technology had more positive learning attitudes and higher achievements than students from the control group. Compared with fully online learning, students of blended courses are more satisfied with the instructional guidance during learning (Lim, Morris, & Kupritz, 2006). It positively impacts students’ comprehension and understanding of the learning content (Woo et al., 2008) and students believe that blended courses enable them to gather knowledge from multiple sources (Bluic,
Ellis, Goodyear, & Piggott, 2011; Orton-Johnson, 2009). Furthermore, students can benefit from increased time and spatial flexibility, wider and easier access to learning sources and a higher level of autonomy in regulating learning (Ashton & Elliot, 2007; Chen & Liu, 2008; Howlett et al., 2011; Owston et al., 2013; Preston et al., 2010).

Despite these overall positive results, some researchers are more cautious and nuanced about this innovation. First, Chong, Tosukhowong and Sakauchi’s (2002) and Chen and Liu (2008) pointed to the technical problems which can occur during web-based delivery of content. But more important, von Konsky et al. (2009) reported that the effectiveness of web-based lectures was not uniform for all students. Concerning this issue, research of López-Pérez, Pérez-López and Rodriguez-Ariza (2011) and Owston et al. (2013) found that compared with lower achieving students, high achievers were the most satisfied with the blended learning approach and it is questioned whether blended learning is as suitable for low achievers as they may not have the independent study skills that blended learning demands.

Moreover, it can be noticed that most literature discusses web-based lecture technology in general without specifying the intended purposes when introducing web-based lectures in education. However, based on the principle of Constructive Alignment, a term coined by Biggs (1996), we believe that it is important to reflect what a teacher want from his/her course and how to provide appropriate teaching and learning activities to reach the objectives. Next to Biggs, also other authors have used this term and it was Fink (2003) who visualized constructive alignment as the triangular model depicted in Figure 1. The basic premise is that courses need to be designed so that the learning activities and assessment tasks are aligned with the learning objectives that are intended in the course. Marinissen & Gratama Van Andel (2012) building on these key components presented different didactic applications of web-based lectures, and showed that they can be used in various contexts, depending on the educational goals lecturers want to reach with their students.

![Figure 1. The key components of Constructive Alignment (Biggs, 1996) visualized by Fink (2003)](image)

The goal of this paper is to investigate and compare students’ perceptions about the benefits of the web-based lectures implemented with different didactical intended purposes and resulting in different application scenarios. Moreover, it is questioned to what extent these perceptions are influenced by students’ personal characteristics and how this interacts with the application scenarios. To this end, web-based lectures were designed and implemented complementary to face-to-face activities. Although the web-based lectures were technically identical, they differed regarding the pedagogical intended purpose and resulted in three application scenarios: (1) web-based lectures as a means of preparation for the face-to-face practical exercises, (2) as a means of repetition of parts of the course, and (3) as a means of extension to supplement classroom lectures.

**Research questions**

The following research questions were investigated regarding the three case studies:

- What are students’ expectations about web-based lectures?
- How do students evaluate the quality of the web-based lectures and how frequently do they watch them?
- What are students’ perceptions concerning the benefits for learning after using the web-based lectures and how does these differ from their pre-experience expectations?
- Are students’ evaluations of web-based lectures affected by their individual characteristics (i.e., gender, learning style, and learning strategy)?
Methods

Context and participants

This study was conducted in the context of an educational innovation project involving three courses taught at the Faculty of Sciences at Ghent University (see Figure 1). The participants were 427 students in the Faculty who were taking one of the following three courses: “Geographical Information Systems” \((N = 88\) students, 53% male students), “Molecular Biology 1” \((N = 76\) students, 47% male students), and “Mathematics 1” \((N = 263\) students, 57% male students). In the Geography course, web-based lectures were used as a means of preparation for practical exercises, in “Molecular Biology 1” they were used as a means of repetition of parts of the course and in “Mathematics 1” as a means of an extension of the course. All web-based lectures were slide-based presentations recorded by means of “Camtasia Studio” screen recording software. As depicted in Figure 2 the software simultaneously captures the video image and audio of the lecturer along with his or her electronic slide presentation. The web-based lectures were subsequently provided via Minerva, the electronic learning environment for students of Ghent University.

![Figure 2](image)

**Figure 2.** Screenshots of the web-based lectures captured by means of “Camtasia Studio” screen recording software and provided to students. From left to right: Geographical Information System (GIS), Molecular Biology 1, and Mathematics 1

Design and procedure

Since students on the different courses were each provided with web-based lectures for different application scenarios, as shown in Figure 3, this study is conducted as a case study design to investigate students’ expectations by means of an intake questionnaire (pre-test) and students’ perceived benefits for learning by means of an exit questionnaire (post-test). Due to the ethical prohibition against withholding benefits from one group of participants (Robson, 2002), a controlled design with random assignment of students within each case was not possible.

![Figure 3](image)

**Figure 3.** The pre- and post-test design to compare web-based lectures with different application scenarios in authentic higher education

Web-based lectures as preparation - Geographical Information System (GIS)

The second-year “Geographic information system (GIS)” course in the BSc in Geography comprised theoretical lectures and seminars (coached exercises) during which students \((N = 88)\) had to work on assignments using GIS-
software. Since the teacher experienced to lose a lot of time to get to learn how to work with the GIS software program, web-based lectures were introduced to students as a way of giving them the opportunity to prepare themselves for the exercises in the class sessions, so that there would be more time to actually work on the exercises in the class sessions. Twenty introductory lessons to the exercises (max. 15 minutes each), explaining features of the software used in the seminars, were recorded as web-based lectures in a studio by the university’s Institute for Permanent Development using Camtasia software, with a teaching assistant serving as lecturer. Students were encouraged by the teaching assistant to prepare themselves for the seminars by viewing these web-based lectures beforehand. In addition, students could use headsets during the seminars, so that they could replay the web-based lectures at will. In this way, students could do the exercises at their own pace, replaying if things were unclear. The teaching assistant was present in the classroom to answer questions. The web-based lectures used in this condition were thus an integral part of the course.

Students were invited to participate in the study and to fill out the intake and exit questionnaires during class sessions. In total 40 students completed both questionnaires which resulted in a response rate of 45%.

**Web-based lectures as repetition - Molecular Biology 1**

The “Molecular Biology 1” course comprised theoretical lectures in molecular biology. Based on previous examinations, it was found that some topics were more difficult than others resulting in low examination scores. Based on this finding, three course topics have been selected for recording, that is “the cis-trans concept,” “the Wobble effect,” and “the alpha-complementation.” These web-based lectures gave students the opportunity to re-listen to these difficult parts of the course at home. The lecturer was the course professor. The lectures were recorded in a studio using Camtasia software. They ranged from 15 to 25 minutes and were provided to 76 students of the second year BSc in Biochemistry and Biotechnology. All students had the same prior knowledge concerning the topics covered. The web-based lectures were available to the students as soon as the topic had been covered in the face-to-face lecture. In this case, then, the web-based lectures were a duplication of parts of the lectures and served as repetition. Additional material to enhance the understanding of the course included scientific articles and movies on selected topics (e.g., DNA replication).

Because the web-based lectures had already been provided to the students at the start of this study, a pre-questionnaire was not administered (student expectations were irrelevant). However, at the end of the course students were invited to participate in the study by filling in the post-questionnaire using the url link. To increase the response rate, students who had not done so within one week were encouraged to fill in the questionnaire on paper during a lecture. In total 51 students completed the post-course questionnaire which resulted in a response rate of 67%.

**Web-based lectures as extension – Mathematics 1**

The “Mathematics 1” course included both theoretical lectures and seminars (coached exercises). Students could prepare themselves for these seminars by making exercises at home. These exercises were placed on Minerva a week in advance, so students had one week to complete them. After the seminar, during which the practical assistant made the exercises on the black board and discussed them thoroughly, the solutions were placed on Minerva as well. Yet, since the teacher indicated that the students group has over the years evolved into a very diverse group with regard to background meaning that some students had had more than 6 hours while other students only have had 3 hours mathematics during secondary education, web-based lectures were additionally introduced to make remediation possible. Web-based lectures were designed as supplementary material to the seminars. Four weeks prior to the end-of-course examination, 65 web-based mini lectures, ranging from 5 to 10 minutes, were provided online to 263 first-year BSc students in Biology, Geology, Biochemistry, Biotechnology, and Geography. In these web-based lectures, the solutions to questions of the last two years' exam were given, in preparation for the exam to come. A teaching assistant, who served as lecturer, worked through the problems on the whiteboard and explained the steps to come to the solutions (verbally). The web-based lectures were recorded in a studio by the Institute for Permanent Development of Ghent University using Camtasia and Mimio technology. Here, then, the web-based lectures served as an extension of the course material.
Compared to the student population of the other two courses, the students following the “Mathematics 1” course formed a very heterogeneous group. Their prior secondary-level mathematics education ranged from two to eight hours a week. In their last year of secondary education, 32% had had 2, 3 or 4 hours of mathematics per week while 56% had had 6, 7 or 8 hours. This learner characteristic was included in the analysis as an extra variable to answer the fourth research question.

Students were invited to participate in the study and to fill in the pre- and post questionnaire online. One week later, to increase the response rate, students who had not filled in the online questionnaire were encouraged to do so on paper during the class session. In total 135 students completed the two questionnaires which resulted in a response rate of 51%.

**Measurements and analysis**

As shown in Table 1, the pre- and post questionnaires consisted of several sections based partly on existing scales. They were available to complete online through the LimeSurvey software tool, which students accessed via a web link sent in an email invitation, or on paper during face-to-face sessions.

<table>
<thead>
<tr>
<th>Table 1, Structure of the pre- and posttest questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pretest</strong></td>
</tr>
<tr>
<td>• Students’ background characteristics</td>
</tr>
<tr>
<td>• Students’ learning strategy and motivation (based on MSLQ)</td>
</tr>
<tr>
<td>• Students’ learning style (VARK)</td>
</tr>
<tr>
<td>• Students’ prior expectations about web-based lectures</td>
</tr>
</tbody>
</table>

Students’ background characteristics comprised gender, academic status (first or second year of undergraduate studies) and students’ prior mathematics education in the final year of secondary education. Students’ learning strategy and motivation was subsequently measured by means of three selected subscales from the Motivated Strategies for Learning Questionnaire (MSLQ) of Pintrich, Smith, Garcia, and McKeachie (1991), which has been successfully utilized in previous educational research (Opdecam, Everaert, Van Keer, & Buysschaert, 2014; Schunk, 2005). The original version of the MSLQ is in English, but we used the Dutch items translated from the original questionnaire using the back translation method of Opdecam et al. (2014). The first selected motivational subscale included in this study is *extrinsic goal orientation* to measure the extent to which students focus on grades and approval from others. Second, students’ *control of learning beliefs* was measured and third, students’ *self-efficacy for learning and performance*, to gauge how students expected and believed that they would perform regarding this particular course. (See Appendix A for the three subscales with the accompanying items and Chronbach’s alphas). The scales were rated on a 7-point Likert scale (1 completely disagree – 7 completely agree.)

Next, to measure student’s learning style, we used the 16-item VARK Questionnaire (Fleming & Mills, 1992). VARK stands for Visual, Aural, Read/Write and Kinesthetic and pertains to students’ different styles of learning. For 16 different provided situations students need to choose the answer which best explains their learning preference; they can tick more than one if a single answer does not match their perception. Consequently, the VARK questionnaire provides users with a profile of their learning preferences and told us something about the way our students tended to take-in and give-out information.

The post-questionnaire gauged students’ use and experiences of the provided web-based lectures and more particularly their perceptions of its benefits for learning and the organization of learning. The different items can be found in the tables in the results section. The post-questionnaire ended with an open answer question to generate more general comments and feedback.

In addition to the primary data from the students’ perspective, we also administered a small lecturer questionnaire concerning their expectations and experiences of the web-based lectures in order to generate feedback to inform future applications (Appendix B). The results of this are integrated into the discussion below.
Results

Web-based lectures as preparation

Expectations

First of all, the expectations of students about using web-based lectures in courses were investigated. Based on a one-sample t-test, the data revealed whether they were significantly higher than the neutral score of “4” (on a 7-point Likert scale). As shown in Table 2, the conclusion can be made that students had high expectations about the benefits of web-based lectures. They believed that web-based lectures would be useful for learning independent from time and geographical location, would help them to prepare for exams and, as preparation for classes, would help them to better understand the content and to achieve better results. All these variables are significantly higher than the neutral score. However students did not expect that their motivation would be improved by using web-based lectures.

Table 2. Students’ expectations of benefits of web-based lectures: Pretest results scored on 7-point Likert scale (N = 40)

<table>
<thead>
<tr>
<th>Web-based lectures…</th>
<th>M (SD)</th>
<th>t (39)</th>
<th>p-value (test value = 4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>… will lead to better understanding of the content.</td>
<td>4.90 (1.37)</td>
<td>4.14</td>
<td>.00</td>
</tr>
<tr>
<td>… will help me to prepare for exams.</td>
<td>5.25 (1.32)</td>
<td>6.01</td>
<td>.00</td>
</tr>
<tr>
<td>… makes learning independent from time and geographical location.</td>
<td>5.45 (1.08)</td>
<td>8.45</td>
<td>.00</td>
</tr>
<tr>
<td>… will help me to achieve better results.</td>
<td>4.70 (1.26)</td>
<td>3.50</td>
<td>.00</td>
</tr>
<tr>
<td>… will increase my motivation for the course.</td>
<td>4.23 (1.51)</td>
<td>.94</td>
<td>.35</td>
</tr>
<tr>
<td>… have an added value in higher education.</td>
<td>5.28 (1.04)</td>
<td>7.77</td>
<td>.00</td>
</tr>
</tbody>
</table>

Experiences

Regarding students’ experiences with the web-based lectures provided (see Table 3), it was found that students had no technical problems while using them. In addition, students reported good image quality and sound quality. Descriptives indicate that 95% of the students in this condition had watched the web-based lectures more than once; the mean (M = 4.83, SD = .68) is also very high (on a 5-point Likert scale). Comments were also positive: “I watched the web lecture because the content was well explained with illustrations,” “I watched the web-based lectures because I want to understand the course,” “Web-based lectures give us important information,” and “Everything was explained well.”

Table 3. Students’ experiences with web-based lectures: Posttest results scored on 7-point Likert scale (N = 40)

<table>
<thead>
<tr>
<th>The provided web-based lectures…</th>
<th>M (SD)</th>
<th>t (39)</th>
<th>p-value (test value = 4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>… had good picture quality.</td>
<td>6.05 (1.31)</td>
<td>11.40</td>
<td>.00</td>
</tr>
<tr>
<td>… had good sound quality.</td>
<td>5.90 (1.15)</td>
<td>10.45</td>
<td>.00</td>
</tr>
<tr>
<td>… were easy to watch on the Internet, without technical problems.</td>
<td>6.03 (1.25)</td>
<td>10.24</td>
<td>.00</td>
</tr>
</tbody>
</table>

Perceived benefits for learning

In this section, a comparison between these high expectations and students’ evaluations after using web-based lectures as preparation of lessons is made. To investigate this, paired sample t-tests were conducted. Results (see Table 4) show that students were more positive than their original expectations with regard to the statement “Web-based lectures will help me to better understand the content.” Students were also more positive in their evaluation about the possibility of learning independently from time and geographical location.
Next, based on a one sample t-test (see Table 5), students in this condition indicate that using web-based lectures have an added value for revising course content, opens up the accessibility for students with special needs, and can help to deepen knowledge.

### Table 5. Students’ overall appreciation of added value of web-based lectures: Posttest results scored on 7-point Likert scale (N = 40)

<table>
<thead>
<tr>
<th>The provided web-based lectures…</th>
<th>M (SD)</th>
<th>t (39)</th>
<th>p-value (test value = 4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>… are an added value for making revision possible.</td>
<td>6.05 (1.20)</td>
<td>10.93</td>
<td>.00</td>
</tr>
<tr>
<td>… can open up the accessibility for students with special needs.</td>
<td>5.75 (1.19)</td>
<td>9.28</td>
<td>.00</td>
</tr>
<tr>
<td>… can help students to deepen knowledge</td>
<td>5.29 (1.52)</td>
<td>5.31</td>
<td>.00</td>
</tr>
</tbody>
</table>

**Impact of students’ characteristics**

As an answer to the 4th research question, and based on a Multiple ANOVA analysis (MANOVA), no significant effects of *individual characteristics* (i.e., gender, learning style, and learning strategy) on students’ evaluation of web-based lectures as preparation (p > .05) were found (see Table 6). In addition, but in contrast with this finding, the interviewed lecturer of this course stated that web-based lectures make students responsible for their own learning process, which turns out to be positive for some students, while other perform worse.

### Table 6. Impact of students’ characteristics: Results of Multiple ANOVA analysis

<table>
<thead>
<tr>
<th>Impact of student characteristics</th>
<th>Wilks’ Lambda</th>
<th>F (10,25)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>.72</td>
<td>.94</td>
<td>.51</td>
</tr>
<tr>
<td>Learning style</td>
<td>.71</td>
<td>1.00</td>
<td>.47</td>
</tr>
<tr>
<td>Extrinsic goal orientation</td>
<td>.80</td>
<td>.65</td>
<td>.76</td>
</tr>
<tr>
<td>Self-efficacy for learning and performance</td>
<td>.60</td>
<td>1.64</td>
<td>.15</td>
</tr>
<tr>
<td>Control of learning beliefs</td>
<td>.70</td>
<td>1.07</td>
<td>.42</td>
</tr>
</tbody>
</table>

**Web-based lectures as repetition**

Since this group had already started the course at the start of the survey, we can only focus on research question 2 to 4.

**Experiences**

When focusing on the *evaluation* of their experience with the web-based lectures, significant positive results can be noticed. Students confirmed that using web-based lectures as repetition helped them to better understand the content of the course (M = 5.38, SD = 1.58, t(31) = 5.62, p < .001) and they agreed that it could help them to be better prepared for exams (M = 5.68, SD = 1.48, t(30) = 6.01, p < .001). Students also confirmed the benefit of studying independently from time and geographical location (M = 4.38, SD = 1.84, t(30) = 3.19, p < .05). No significant effects were found regarding the perceived learning outcomes (M = 4.32, SD = 1.68, t(30) = 1.07, p > .05) nor
students perceived web-based lectures as a means to increase the motivation for the courses \( M = 3.71, SD = 1.49, r(30) = -1.09, p > .05 \).

Students had no technical problems using web-based lectures \( M = 5.10, SD = 1.58, t(30) = 8.42, p < .001 \) and reported good image quality \( M = 5.87, SD = 1.20, t(30) = 3.97, p < .001 \) and sound quality \( M = 5.81, SD = 1.20, t(30) = 8.65, p < .001 \). Means of the frequency of watching the web-based lectures in this format were low \( M = 2.63, SD = 1.66 \), with 53% of the students not watching them more than once. When asking for the reasons behind this low score, students reported: “no time,” “the web-based lectures are not necessary,” “superfluous.” However, a minority reported contradictory evaluations: “the movies were clearer than the other course materials,” “they visualized the content,” and “they have an added value for preparation of exams.” Moreover, after using the web-based lectures, students in this course agreed with the statement that using web-based lectures was an added value for making revision possible \( M = 5.29, SD = 1.46, t(40) = 5.33, p < .001 \), opens up accessibility for students with special needs \( M = 5.06, SD = 1.52, t(40) = 4.43, p < .001 \) and can help to deepen knowledge \( M = 5.07, SD = 1.57, t(40) = 4.37, p < .001 \). In addition, the interviewed lecturer of this course agrees with the statement that using web-based lectures in this way is an effective tool for increasing understanding of the course. He was very positive about this experience with the web-based lecture and professed an interest in making more recordings in the future.

Impact of students’ characteristics

Because of the lack of data on learning style and learning strategy, only gender could be included in this part of the analysis. Results indicate no relation between learning characteristics and using web-based lectures as repetition (Wilks’ Lambda = .72, \( F(13, 17) = .52, p > .05 \)).

Web-based lectures as extension

Expectations

Students’ expectations about using web-based lectures in this condition were also higher than the neutral “4”, using one sample t-tests. These data indicate that students agreed that using web-based lectures as an extension of the course would be an added value \( M = 5.10, SD = 1.08, t(130) = 1.65, p < .001 \) and that they would be better prepared for exams \( M = 5.37, SD = .99, t(130) = 15.93, p < .001 \). They also agreed that using web-based lectures would be useful for better understanding course content \( M = 5.00, SD = 1.03, t(130) = 11.11, p < .001 \), for learning independently of time and geographical location \( M = 4.70, SD = 1.14, t(130) = 9.71, p < .001 \)) and for achieving better results \( M = 4.78, SD = 1.09, t(130) = 8.17, p < .001 \). All these variables were significantly higher than the neutral score of “4” \( p < .001 \). On the other hand, students did not expect web-based lectures to improve their motivation to learn \( M = 4.20, SD = 1.22, t(130) = 1.86, p > .05 \).

Experiences

When using web-based lectures for this purpose, students experienced no technical problems \( M = 5.33, SD = 1.15, t(116) = 10.76, p < .001 \) and reported strong picture quality \( M = 5.32, SD = 1.17, t(116) = 12.60, p < .001 \) and sound quality \( M = 5.18, SD = 1.87, t(116) = 12.30, p < .001 \). Only 20% of the students in this condition reported watching the web-based lectures more than once and more than 88% of them reported watching half of the web-based lectures or even less. In fact, qualitative data show us that most students report not watching the web-based lectures at all, some because of lack of time. Two students stated that they watched only some parts of the lectures since a 20 minute lecture was found to be too long to watch. Another student indicated that he only watched a part of the lecture out of curiosity.

Perceived Benefits for learning

In contrast to the condition in which web-based lectures were used as course preparation, paired sample t-tests reveal that in this condition expectations before using web-based lectures were significantly higher than the evaluation afterwards (see Table 7).
Table 7. Students’ expectations and evaluations of benefits of web-based lectures: Pretest and posttest results compared (7-point Likert scale; N = 51)

<table>
<thead>
<tr>
<th>Web-based lectures…</th>
<th>Pretest M (SD)</th>
<th>Posttest M (SD)</th>
<th>t(110)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>… has led to better understanding of the content.</td>
<td>5.04 (1.01)</td>
<td>4.07 (1.57)</td>
<td>3.41</td>
<td>.00</td>
</tr>
<tr>
<td>… will help me to prepare for exams.</td>
<td>5.43 (1.00)</td>
<td>4.62 (1.55)</td>
<td>5.44</td>
<td>.00</td>
</tr>
<tr>
<td>… makes learning independent from time and geographical location.</td>
<td>4.97 (1.16)</td>
<td>4.10 (1.49)</td>
<td>6.06</td>
<td>.00</td>
</tr>
<tr>
<td>… will help me to achieve better results.</td>
<td>4.81 (1.10)</td>
<td>3.49 (1.37)</td>
<td>10.66</td>
<td>.00</td>
</tr>
<tr>
<td>… has increased my motivation for the course.</td>
<td>4.22 (1.22)</td>
<td>3.40 (1.38)</td>
<td>5.57</td>
<td>.00</td>
</tr>
<tr>
<td>… has an added value in higher education.</td>
<td>5.10 (1.09)</td>
<td>4.76 (1.49)</td>
<td>2.47</td>
<td>.02</td>
</tr>
</tbody>
</table>

However, these results should be nuanced. When students on this course were asked for a final evaluation of using web-based lectures, results are more positive. These students report that web-based lectures are an added value for making revision possible (M = 5.07, SD = 1.28, t(126) = 9.43, p < .001) and improve accessibility for students with special needs (M = 5.31, SD = 1.21, t(126) = 12.22, p < .001). In addition, the interviewed lecturer of this course agrees with the statement that web-lectures support students to prepare exams more effective. Compared with these high evaluation scores, students produced lower scores on the question of whether web-based lectures can help to deepen knowledge (M = 4.78, SD = 1.39, t(126) = 6.34, p < .001).

Impact of students’ characteristics

In a final step, significant effects of self-efficacy (negative effect) (Wilks’ Lambda = .80, F(10, 94) = 2.39, p < .05) and expectancy (Wilks’ Lambda = .80, F(10, 94) = 2.36, p < .05) on the evaluation of web-based lectures as an extension were found. This means that students who did not themselves believe that they would pass the course and students who scored high on control of learning beliefs have a significantly higher change to perceive web-based lectures as an added value. Moreover, it was found that students’ ability profile / prior knowledge (operationalized by the number of hours of mathematics received in secondary education) had an effect on the appreciation of the provision of web-based lectures as an extension. Students who had had less than 6 hours of mathematics per week in secondary education reported that using web-based lectures helped them to better understand the subject matter to a greater extent than did students who had received 6 or 8 hours weekly (M_{6<h} = 4.49, M_{6,8} = 3.77, t(112) = -2.40, p < .05).

The former group also found web-based lectures more useful for preparing for exams (M_{6<h} = 4.91) than did the latter group (M_{6,8} = 4.35, t(112) = -.871, p =.064 (marginally significant).

Discussion and conclusion

The goal of this study was to investigate and compare students’ perceptions about the benefits of the web-based lectures with different didactical intended purposes and resulting in different application scenarios. Moreover, prior to the research reported in this paper, little specific research such as López-Pérez et al. (2011), Owston et al. (2013) and von Konsky et al. (2009) has been done to study the impact of individual student characteristics on the evaluation of different scenarios of applying web-based lectures. Our results indicate that overall students evaluate this kind of technology as an added value. This confirms previous research which has pointed to the general benefits of using web-based lectures to complement face-to-face courses (Day, 2008; Day & Foley, 2006; Howlett et al., 2011; Lim & Morris, 2009; Owston et al., 2013; Taradi et al., 2005). In particular, research shows that web-based lectures are seen as an added value for making revision possible (Lai et al., 2011), help students to learn independent from time and geographical location (Ashton & Elliot, 2007; Chen & Liu, 2008; Howlett et al., 2011; Owston et al., 2013; Preston et al., 2010) and provide the possibility of helping to deepen knowledge (Woo et al., 2008).

When comparing different scenarios of applying web-based lectures, our results suggest that the web-based lectures used as a means of course preparation is evaluated most positively. Web-based lectures are appreciated the least
when used as a means of course extension. However, regarding the latter, results need to be nuanced by taking into account student characteristics. It has been found that students with a poor background and low self-efficacy for mathematics and an internal locus of control have a significantly higher appreciation for the application of web-based lectures as a means of extension. This means that web-based lectures as a means of extension are especially beneficial for low-achieving students. This was however not the case when web-based lectures were implemented as a means of preparation since the lecturer of the GIS course noticed that while web-based lectures enhanced flexibility for the student and made them responsible for their own learning process, this feature was positive for some students but negative for others. This is consistent with the concerns of López-Pérez et al. (2011) and Owston et al. (2013) that web-based lectures would not be suitable for low-achievers as they may not have the independent study skills that blended learning demands. These results support the indications of von Konsky et al. (2009) and stress the need for further research regarding the impact of individual characteristics regarding different scenarios of applying web-based lectures.

Based on these results, the following conclusions with respect to the application of web-based lectures in education can be drawn. First of all, interviews with the course lecturers in our study revealed that perceiving the integration of web-based lectures as an added value depends on the intended purposes of the course. Every application of web-based lectures can be an added value, as long as the desired educational goal is taken into account which is consistent with the concept of “constructive alignment” (Biggs, 1996). Using web-based lectures as a repetition is most beneficial when lecturers want to provide the opportunity for students to repeat difficult lectures and concepts. Using them as preparation is most beneficial for lecturers who want to save time by replacing basic lessons with recordings for students to watch at home and to provide more time for exercises and the possibility of answering questions during the limited time available for face-to-face lessons. Finally, using web-based lectures as an extension of the course can be most beneficial for students who need more exercises or for those who want a deepening of the course content. Secondly, student characteristics can play a significant role and should be taken into account when introducing web-based lectures. In this respect, web-based lectures are a way to differentiate and meet the needs of all students in higher education. Thirdly, and in contrast to the findings of Chen & Liu (2008), students in all courses in our study reported to experience no technical problems while using web-based lectures. Once lectures have access to, or purchased, screen recording software (e.g., “Camtasia Studio”), they can easily start creating web-based lectures. Hence, using this technology does not imply the need for lecturers and students to learn difficult technical skills before they can use this technology for educational activities. Concluding, the potential of web-based lectures lies in its adaptability for use in several scenarios according to the needs of students and the didactical intended purposes.

Limitations and further research

A limitation of our study is the absence of a control group. Another limitation was the absence of pre-test data in the group where web-based lectures were used as a repetition, because the course had already begun at the start of the survey. Therefore, no clear conclusion about learning style could be made.

In this study we measured students’ satisfaction and perceived educational effects of the application of web-based lectures from different scenarios, in future research it would be interesting to study the effects on student achievement as well.

References


Appendix A

Subscales with corresponding items based on the original MSLQ
(1 completely disagree – 7 completely agree)

<table>
<thead>
<tr>
<th>Cronbach’s alpha</th>
<th>Extrinsic goal orientation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1. Getting a good grade in this class is very important for me.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. If I can, I want to get better grades in this class than most of the other students.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. I want to do well in this class because it is important to show my ability to my family and friends</td>
<td></td>
</tr>
<tr>
<td>0.74</td>
<td>Control of learning Beliefs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1. If I study in appropriate ways, then I will be able to learn the material in this course.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. It is my own fault if I don’t succeed in this course.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. If I try hard enough, then I will understand the course material.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. If I don’t understand the course material, it is because I didn’t try hard enough.</td>
<td></td>
</tr>
<tr>
<td>0.74</td>
<td>Self-Efficacy for Learning &amp; Performance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1. I believe I will receive an excellent grade in this class.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. I’m certain I can understand the most difficult material presented in the readings for the course.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. I’m confident I can learn the basic concepts taught in this course.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. I’m confident I can understand the most complex material presented by the lecturer in this course.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5. I’m confident I can do an excellent job on the exam of this course.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6. I expect to do well in this class.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7. I’m certain I can master the skills being taught in this class.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8. Considering the difficulty of this course, the teacher, and my skills, I think I will do well in this class.</td>
<td></td>
</tr>
<tr>
<td>0.94</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Appendix B

Questionnaire for the lecturers

General questions

1. Name and course where you implemented web lectures:
2. What was the purpose of the application of web lectures? What were your expectations?
3. Did the application of web lectures meet your expectations? Why/Why not?
4. Did any technical or organizational problems arise? Explain.
5. Would you use web lectures again in the future? Why/Why not?
6. What aspects can be improved?

Give your opinion on the following statements

5. The use of web lectures is an effective aid to enhance the understanding of the course.
6. Students perform better using the web lectures.
A Pilot Meta-Analysis of Computer-Based Scaffolding in STEM Education

Brian R. Belland¹, Andrew E. Walker¹, Megan Whitney Olsen¹ and Heather Leary²

¹Department of Instructional Technology and Learning Sciences, Utah State University, USA // ²Institute of Cognitive Science, University of Colorado-Boulder, USA // brian.belland@usu.edu // andy.walker@usu.edu // m.whitney.olsen@gmail.com // heatherleary@gmail.com

*Corresponding author

(Submitted October 8, 2013; Revised March 15, 2014; Accepted April 10, 2014)

ABSTRACT

This paper employs meta-analysis to determine the influence of computer-based scaffolding characteristics and study and test score quality on cognitive outcomes in science, technology, engineering, and mathematics education at the secondary, college, graduate, and adult levels. Results indicate that (a) computer-based scaffolding positively influences learning (g = 0.53), (b) studies with zero threats to internal validity had lower effect sizes than studies with two threats, (c) studies with one threat to external validity had higher effect sizes than studies with zero threats, (d) studies with no fading had higher effect sizes than studies with fixed fading, and (e) students performed better when using conceptual scaffolds than with metacognitive scaffolds. There were no differences based on study design, generic vs. specific, paired intervention, assessment level, or intended learning outcome. Meta-regression indicated that fading or lack thereof explained 30% of the variability in outcomes. The significance of this paper lies in its potential to steer scaffold designers away from fixed fading and metacognitive scaffolds, and toward studying scaffolding in authentic contexts, rather than laboratories. Furthermore, this study indicates that a more comprehensive scaffolding meta-analysis is warranted.

Keywords

Computer-based scaffolding, Meta-analysis, Assessment quality, Methodological quality, STEM education

A Pilot meta-analysis of computer-based scaffolding in stem education

Despite much primary research and review work about scaffolding (Kali & Linn, 2008), scaffolding meta-analyses remain an emergent line of inquiry. Meta-analyses have been published on specific types of scaffolding including dynamic assessment (Swanson & Lussier, 2001), scaffolding for students with learning disabilities (Swanson & Deshler, 2003), and scaffolding in multimedia instruction (Lin, Ching, Ke, & Dwyer, 2007). However, a more comprehensive examination of scaffolding is needed to inform researchers and designers of the most effective characteristics of and approaches to researching scaffolding. In this paper, we use meta-analysis to determine the influence of computer-based scaffolding characteristics, study and test score quality, and assessment level on cognitive outcomes in science, technology, engineering, and mathematics (STEM) education. This paper is significant in that it (a) provides evidence of the effectiveness of scaffolding, (b) can guide future research on scaffolding, and (c) provides for data-driven scaffolding design decisions.

Theoretical framework

Computer-based scaffolding

Definition

As Vygotsky (1962) noted, “The only good kind of instruction is that which marches ahead of development and leads it” (p. 104). Scaffolding can facilitate such instruction by providing conceptual, procedural, strategic, and metacognitive support that bridges the gap between what students can do on their own and what they can do with the help of a more capable other (Hannafin, Land, & Oliver, 1999; Wood, Bruner, & Ross, 1976). As originally defined, scaffolding referred to dynamic support provided by a teacher or other more capable other that enabled children to solve problems (Wood et al., 1976). The emergence of personal computers has allowed computer-based scaffolding to be developed to supplement teacher scaffolding (Hannafin et al., 1999; Saye & Brush, 2002). Computer-based scaffolding (hereafter referred to as “scaffolding”) can promote success in rich problem solving contexts that require students to go beyond filling out worksheets or listening to a teacher lecture.
Meta-analyses of specific scaffolding types (scaffolding for dynamic assessment, learning disabilities, and multimedia instruction) indicate that scaffolding-related interventions were associated with increased student learning (Lin et al., 2007; Swanson & Deshler, 2003; Swanson & Lussier, 2001). Dynamic assessment led to an average effect size of 0.96, and there was a similarly positive relationship between explicit practice and learning (Swanson & Lussier, 2001). Scaffolding in multimedia instruction led to an average effect size of 0.02 (Lin et al., 2007). Thus, a meta-analysis that covers a wider range of scaffolding types is warranted.

**Scaffolding characteristics**

Scaffolding can (a) enlist interest in the target task (Wood et al., 1976), (b) maintain direction (Wood et al., 1976), (c) reduce complexity (Reiser, 2004), (d) highlight important problem features (Reiser, 2004), (e) help students manage frustration (Wood et al., 1976), (f) model expert processes (van de Pol, Volman, & Beishuizen, 2010), and (g) elicit articulation (Reiser, 2004). Scaffolding can do this through such strategies as making thinking visible (Kali & Linn, 2008). No scaffolding serves all these aims, and no literature compares such strategies to indicate which are most effective in which contexts.

Though scaffolding originally referred to support for children’s problem solving abilities (Wood et al., 1976), recent scaffolding work also supports (a) other higher-order thinking skills (e.g., argumentation and evaluation), and (b) knowledge integration - the ability to “expand, revise, restructure, reconnect and reprioritize” scientific models (Linn, 2000, p. 783). The literature does not indicate which scaffolding type leads to stronger outcomes.

Scaffolding is not a stand-alone intervention. Rather, it is used along with instructional approaches (paired interventions) such as problem-based learning (PBL) (Saye & Brush, 2002), learning by design (Puntambekar & Kolodner, 2005), and case-based learning (Lajoie, Lavigne, Guerrera, & Munsie, 2001). No literature comprehensively compares the effectiveness of scaffolding when used with these different approaches.

Many debate whether scaffolding should be generic or context-specific. For example, one finding indicated that generic scaffolds promoted deeper reflection among middle school students than content-specific scaffolds (Davis, 2003), while another finding indicated that content-specific scaffolds were more effective when teachers provided one-to-one scaffolding related to a general argumentation framework (McNeill & Krajcik, 2009).

Many researchers argue that scaffolds must be faded as students gain skill to promote transfer of responsibility (Pea, 2004; Puntambekar & Hübscher, 2005). In one-to-one scaffolding, this is done through continual diagnosis of student performance (van de Pol et al., 2010). The only two ways this has been accomplished among computer-based scaffolds to support ill-structured problem solving are making the scaffolds disappear (a) on a fixed schedule (Li & Lim, 2008) or (b) when students indicate that they do not need the scaffolding anymore (Metcalf, 1999). It is not clear if such approaches lead to greater learning or transfer of responsibility.

**Scaffolding and methodological quality**

Methodological quality considerations include threats to internal validity and external validity (Gall, Gall, & Borg, 2003; Shadish & Myers, 2001), as well as test score reliability and validity (Messick, 1989). It may be unrealistic to expect that a scaffolding study have zero threats to validity due to the need to study scaffolding in contexts in which students can collaboratively solve authentic problems - contexts that do not include laboratories. In the three existing scaffolding meta-analyses, as methodological quality decreased, effect size magnitude increased (Lin et al., 2007; Swanson & Deshler, 2003; Swanson & Lussier, 2001). This highlights the need to code for methodological quality in a wider scaffolding meta-analysis.

**Scaffolding and assessment level**

When evaluating scaffolding's effectiveness, one needs to consider assessment level (e.g., concept, principles, and application) (Messick, 1989; Sugrue, 1995). For example, a fact-based test may not be the best vehicle to evaluate scaffolding designed to promote problem solving ability.
Remaining questions

Scaffolding guidelines exist in such areas as online discussion (Choi, Land, & Turgeon, 2005), science inquiry (Linn, 2000; Quintana et al., 2004), problem solving (Ge & Land, 2004; Kolodner, Owensby, & Guzdial, 2004), and argumentation (Belland, Glazewski, & Richardson, 2008; Jonassen & Kim, 2010). Authors often gather empirical support for guidelines (Belland, 2010; Ge & Land, 2003), but the sheer volume of scaffolding frameworks and conflicting advice leads one to desire a comprehensive assessment of scaffolding. Furthermore, the effect of a specific feature inspired by a guideline is rarely isolated in empirical studies. The lack of clarity is not due to a lack of research, as there is a staggering volume of empirical research on scaffolding. Individual empirical studies are the engine that drives educational research, but in accumulating thousands of studies on scaffolding, researchers run the risk of “knowing less than we have proven” (Glass, 1976, p. 8). Meta-analysis is one way to avoid this risk (Chambers, 2004; Glass, 1976).

Research questions

In this preliminary meta-analysis, we address the following research questions:

1. To what extent do scaffolding characteristics (strategy, intended outcome, fading schedule, intervention, and paired intervention) influence cognitive outcomes?
2. To what extent does methodological quality (study design, internal threats to validity, external threats to validity, and test score validity and reliability) influence cognitive outcomes?
3. To what extent does assessment level (concept, principle, application) influence cognitive outcomes?
4. To what extent do the combination of internal threats, external threats, reliability, fading, and scaffolding intervention influence cognitive outcomes?

Method

Inclusion criteria

To be included, studies needed to (a) cover primary, middle level, secondary, college/vocational, graduate/professional, and adult students, (b) compare a scaffolding treatment with a comparison condition (absence of scaffolding), (c) report quantitative, cognitive outcomes (e.g., problem-solving ability, conceptual understanding), and (d) report enough data to support effect size calculation. When more than one source reported the same data (e.g., a dissertation and a journal article), the source with the most detail (e.g., dissertation) was included.

Literature search

The literature search began with a comprehensive review of scaffolding strategies (Kali & Linn, 2008) before proceeding to existing meta-analyses (Lin et al., 2007; Swanson & Deshler, 2003; Swanson & Lussier, 2001). Ninety-four studies were identified as candidates for inclusion on first pass. Upon application of the inclusion criteria by two research team members to each study identified as candidates for inclusion, the 94 studies were reduced to seven, for reasons including (a) not enough information to calculate an effect size, (b) lack of quantitative, cognitive outcomes, (c) lack of a control condition in which students did not receive scaffolding, and (d) the intervention did not meet the definition of computer-based scaffolding.

Coding scheme

Two coders independently coded each study for several characteristics (see Table 1).
Table 1. Coding Scheme

<table>
<thead>
<tr>
<th>Contextual Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paired intervention (e.g., problem-based learning)</td>
</tr>
<tr>
<td>Assessment level – conceptual, principles, application (Sugrue, 1995)</td>
</tr>
<tr>
<td>Education level (e.g., middle level)</td>
</tr>
<tr>
<td>Discipline (e.g., physical science)</td>
</tr>
<tr>
<td>Collection year</td>
</tr>
<tr>
<td>Institution name (of the primary author)</td>
</tr>
<tr>
<td>Attrition treatment and attrition control (% of total assigned to treatment or control conditions)</td>
</tr>
<tr>
<td>Study design (e.g., random, group random, or quasi experimental)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scaffold intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scaffolding strategy (e.g., make thinking visible) (Kali &amp; Linn, 2008)</td>
</tr>
<tr>
<td>Scaffolding function (e.g., reduce complexity) (Wood et al., 1976)</td>
</tr>
<tr>
<td>Scaffolding intervention (e.g., conceptual) (Hannafin et al., 1999)</td>
</tr>
<tr>
<td>Fading (none; human-adapted; fixed; self-selected)</td>
</tr>
<tr>
<td>Generic or specific</td>
</tr>
<tr>
<td>Scaffolding Outcome (higher-order thinking skills or knowledge integration)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Required information for effect size calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Means, SD, or ANOVA or other statistical comparison test values</td>
</tr>
<tr>
<td>N for treatment and control</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Threats to internal validity (Gall et al., 2003)</th>
</tr>
</thead>
<tbody>
<tr>
<td>History</td>
</tr>
<tr>
<td>Maturation</td>
</tr>
<tr>
<td>Testing</td>
</tr>
<tr>
<td>Instrumentation</td>
</tr>
<tr>
<td>Statistical regression</td>
</tr>
<tr>
<td>Differential selection</td>
</tr>
<tr>
<td>Experimental mortality</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Threats to external validity (Shadish &amp; Myers, 2001)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Limited description of treatment</td>
</tr>
<tr>
<td>Multiple treatment interaction</td>
</tr>
<tr>
<td>Experimenter effect</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test Score Quality Reporting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test score validity reporting</td>
</tr>
<tr>
<td>Test score reliability reporting</td>
</tr>
</tbody>
</table>

Note. Bolded coding items indicate a category that was added or heavily modified as a result of pilot coding. Internal validity and external validity threats were coded for the degree to which the threat could account for study results; 0 meant not a plausible threat and 3 meant it could account for all variance in the outcome. Test score quality reporting was assessed as (a) strong, if authors reported validity or reliability data for their own sample, (b) attempt, if authors reported validity or reliability data from other studies, or (c) none, if no validity or reliability data was reported.

Analyses

Given the diversity of research quality, interventions, populations, and sample sizes among existing primary research, effect size estimate precision varied. Thus, a conversion was made from Cohen’s d to Hedges’ g for all outcomes (Cooper, 1989).

To examine potential bias a funnel plot was generated. Figure 1 shows standard error and Hedge’s g for each study. The figure shows a general lack of symmetry suggesting publication bias. To investigate further, a cumulative forest plot (Sutton, 2009) was run. The nine most precise studies \((g = 0.33)\) have a substantially lower estimate than all 17 studies \((g = 0.53)\). This implies bias either due to publication or small study effects. This is somewhat anticipated, as the primary source of studies for this meta-analysis was existing reviews that tend not to cover “grey literature” such as book chapters, conference papers, or dissertations (Borenstein, Hedges, Higgins, & Rothstein, 2009). Only one included study was a dissertation; the rest were peer-reviewed journal articles.
Research questions 1-3 were addressed with a Q-test based on analysis of variance. Pairwise differences were assessed using a Z-test (Borenstein et al., 2009). For research question 4, meta-regression was employed. Since the studies in this analysis draw from fundamentally different populations a random effects model was employed. All significance testing assumed an alpha level of .05. Data analysis was conducted using STATA 11.

Results

Before addressing our research questions, we present overall analyses. Seven studies with 17 outcomes met our inclusion criteria (Bagno & Eylon, 1997; Clement, 1993; Foley, 1999; Mayer, 1989; Nathan, Kintsch, & Young, 1992; Ronen & Eliahu, 2000; White & Frederiksen, 1998). Covered subject areas/grade levels are: high school physics (2 studies), middle school physical science (1 study), college mathematics (1 study), college mechanics (1 study), high school science (1 study), and middle and high school science (1 study). An I-squared test indicates a relatively high level of inconsistency (See Figure 2) across effect size estimates (78.0%, p = 0.001). The overall effect size ($g = 0.53$) is statistically significant, $z = 4.65$, $p = 0.001$, and of a medium magnitude (Cohen, 1988).

Boxed Hedges’ $g$ values are statistically greater than zero, $p < 0.05$. The forest plot shows estimates and confidence levels for each outcome. The diamond is a summary estimate and confidence interval for the overall effect.

Research question 1: To what extent do scaffolding characteristics (strategy, intended outcome, fading, intervention, and paired intervention) influence cognitive outcomes?

Table 2 shows point estimates and 95% confidence intervals for Hedges' $g$ for each coded scaffolding characteristic. Examining strategy, one notices that there was no significant difference in cognitive outcomes between generic and context-specific scaffolds, and that the effect size for each is significantly greater than 0, $p < 0.05$. This suggests that there are no differences in influence on cognitive outcomes between generic and context-specific scaffolds, though caution is warranted as only one coded study included a generic scaffold.

There were no significant differences between scaffolding that supports higher-order skills and scaffolding that supports knowledge integration (see Table 2). This provides preliminary evidence that scaffolding designed to support each intended outcome is equally effective. As the confidence intervals are fairly wide, examining more studies is necessary to have great confidence in no significant differences.
Table 2. Influence of scaffolding characteristics on cognitive outcomes

<table>
<thead>
<tr>
<th>Scaffolding Characteristic</th>
<th>N\textsubscript{studies}</th>
<th>N\textsubscript{outcomes}</th>
<th>g</th>
<th>CI\textsubscript{Lower}</th>
<th>CI\textsubscript{Upper}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generic\textsuperscript{a}</td>
<td>1</td>
<td>1</td>
<td>0.72</td>
<td>0.41</td>
<td>1.03</td>
</tr>
<tr>
<td>Specific\textsuperscript{a}</td>
<td>6</td>
<td>16</td>
<td>0.52</td>
<td>0.28</td>
<td>0.75</td>
</tr>
<tr>
<td>Intended Outcome</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Higher order skills\textsuperscript{a}</td>
<td>2</td>
<td>7</td>
<td>0.44</td>
<td>0.31</td>
<td>0.74</td>
</tr>
<tr>
<td>Knowledge integration\textsuperscript{a}</td>
<td>5</td>
<td>10</td>
<td>0.59</td>
<td>0.27</td>
<td>0.90</td>
</tr>
<tr>
<td>Fading Schedule</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None\textsuperscript{a}</td>
<td>6</td>
<td>10</td>
<td>0.79</td>
<td>0.47</td>
<td>1.11</td>
</tr>
<tr>
<td>Fixed</td>
<td>2</td>
<td>7</td>
<td>0.20</td>
<td>0.00</td>
<td>0.41</td>
</tr>
<tr>
<td>Scaffolding Intervention</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conceptual\textsuperscript{a}</td>
<td>6</td>
<td>12</td>
<td>0.67</td>
<td>0.37</td>
<td>0.97</td>
</tr>
<tr>
<td>Metacognitive</td>
<td>1</td>
<td>5</td>
<td>0.25</td>
<td>0.00</td>
<td>0.50</td>
</tr>
<tr>
<td>Paired Intervention</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inquiry-based learning\textsuperscript{a}</td>
<td>3</td>
<td>12</td>
<td>0.39</td>
<td>0.16</td>
<td>0.61</td>
</tr>
<tr>
<td>Problem solving\textsuperscript{a}</td>
<td>4</td>
<td>5</td>
<td>0.53</td>
<td>0.31</td>
<td>0.75</td>
</tr>
</tbody>
</table>

\textbf{Note.} \textsuperscript{a}Significantly greater than an effect size of 0. p < .05.
Turning to Fading Schedule in Table 2, studies in which scaffolding was not faded had higher effect sizes ($g = 0.79$) than studies that employed fixed fading ($g = 0.20$), $z(16) = 3.02, p = .001$. This implies that fixed fading harms cognitive outcomes (see Figure 3).

<table>
<thead>
<tr>
<th>Fading</th>
<th>studies</th>
<th>outcomes</th>
<th>Hedges’ g</th>
<th>lower</th>
<th>upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
<td>6</td>
<td>10</td>
<td>0.79</td>
<td>0.47</td>
<td>1.11</td>
</tr>
<tr>
<td>fixed</td>
<td>2</td>
<td>7</td>
<td>0.20</td>
<td>0.00</td>
<td>0.41</td>
</tr>
<tr>
<td>Overall</td>
<td>7</td>
<td>17</td>
<td>0.53</td>
<td>0.31</td>
<td>0.75</td>
</tr>
</tbody>
</table>

*Figure 3. Learning by fading*

Note boxed $Hedges’ g$ values are statistically greater than zero, $p < .05$. Diamonds are summary scores, representing Hedges’ $g$ and confidence intervals for two or more outcomes; the final diamond represents all outcomes. “Lower” and “upper” represent the effect sizes’ 95% confidence interval limits.

As indicated in Table 2 and Figure 4, studies using conceptual ($g = 0.67$) scaffolds exhibited superior learning outcomes to those using metacognitive ($g = 0.25$) scaffolds, $z (16) = 3.29, p = .01$. This emergent result should be interpreted with caution as only one study used metacognitive scaffolds.

<table>
<thead>
<tr>
<th>Intervention</th>
<th>studies</th>
<th>outcomes</th>
<th>Hedges’ g</th>
<th>lower</th>
<th>upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>conceptual</td>
<td>5</td>
<td>12</td>
<td>0.67</td>
<td>0.37</td>
<td>0.97</td>
</tr>
<tr>
<td>metacognitive</td>
<td>1</td>
<td>5</td>
<td>0.25</td>
<td>0.00</td>
<td>0.50</td>
</tr>
<tr>
<td>Overall</td>
<td>7</td>
<td>17</td>
<td>0.53</td>
<td>0.31</td>
<td>0.75</td>
</tr>
</tbody>
</table>

*Figure 4. Learning by scaffolding intervention*

Boxed $Hedges’ g$ values are statistically greater than zero, $p < .05$. Diamonds are summary scores, representing Hedges’ $g$ and confidence intervals for two or more outcomes; the final diamond represents all outcomes. “Lower” and “upper” represent limits of the effect sizes’ 95% confidence intervals.

There were no significant differences between studies based on paired intervention: Inquiry-based learning led to an effect size of 0.39 and problem-solving led to an effect size of 0.53 (see Table 2). Of note, scaffolding is used in the context of other instructional approaches, so it is important to examine more studies covering a wider range of paired interventions.
Research question 2: To what extent does methodological quality (study design, internal threats to validity, external threats to validity, and test score validity and reliability) influence cognitive outcomes?

As indicated in Table 3, there were no statistically significant differences between study designs. Results from quasi-experimental designs parallel results of true random experiments, although the confidence intervals would likely shrink with the inclusion of more studies.

<table>
<thead>
<tr>
<th>Methodological Quality Indicator</th>
<th>$N_{studies}$</th>
<th>$N_{outcomes}$</th>
<th>$g$</th>
<th>CI$_{Lower}$</th>
<th>CI$_{Upper}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Study Design</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quasi-experimental$^a$</td>
<td>3</td>
<td>3</td>
<td>0.60</td>
<td>0.16</td>
<td>1.05</td>
</tr>
<tr>
<td>Group random$^b$</td>
<td>2</td>
<td>8</td>
<td>0.43</td>
<td>0.13</td>
<td>0.73</td>
</tr>
<tr>
<td>Random$^a$</td>
<td>2</td>
<td>6</td>
<td>0.66</td>
<td>0.19</td>
<td>1.13</td>
</tr>
<tr>
<td><strong>Validity Reporting</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None$^c$</td>
<td>7</td>
<td>17</td>
<td>0.53</td>
<td>0.30</td>
<td>0.75</td>
</tr>
<tr>
<td>Attempt</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Strong</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Reliability Reporting</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None$^d$</td>
<td>6</td>
<td>11</td>
<td>0.67</td>
<td>0.37</td>
<td>0.97</td>
</tr>
<tr>
<td>Attempt</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Strong$^c$</td>
<td>1</td>
<td>6</td>
<td>0.25</td>
<td>0.01</td>
<td>0.50</td>
</tr>
<tr>
<td><strong>Number of Internal Threats</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>2</td>
<td>5</td>
<td>0.29</td>
<td>-0.11</td>
<td>0.68</td>
</tr>
<tr>
<td>One</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Two$^a$</td>
<td>3</td>
<td>4</td>
<td>0.89</td>
<td>0.67</td>
<td>1.09</td>
</tr>
<tr>
<td>Three$^a$</td>
<td>4</td>
<td>8</td>
<td>0.45</td>
<td>0.16</td>
<td>0.74</td>
</tr>
<tr>
<td><strong>Number of External Threats</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None$^b$</td>
<td>1</td>
<td>5</td>
<td>0.25</td>
<td>0.00</td>
<td>0.50</td>
</tr>
<tr>
<td>One$^b$</td>
<td>2</td>
<td>4</td>
<td>1.00</td>
<td>0.63</td>
<td>1.37</td>
</tr>
<tr>
<td>Two</td>
<td>2</td>
<td>3</td>
<td>0.36</td>
<td>-0.35</td>
<td>1.07</td>
</tr>
<tr>
<td>Three</td>
<td>2</td>
<td>2</td>
<td>0.44</td>
<td>-0.10</td>
<td>0.98</td>
</tr>
<tr>
<td>Four</td>
<td>1</td>
<td>3</td>
<td>0.83</td>
<td>-0.13</td>
<td>1.79</td>
</tr>
</tbody>
</table>

*Note.* $^a$Significantly greater than an effect size of 0, $p < .05$. $^b$Does not add up to 7 because internal threats are associated with individual outcomes. Some studies included more than one outcome. $^c$Does not add up to 7 because external threats are associated with individual outcomes. Some studies included more than one outcome.

None of the studies reported validity data. Thus, no difference could be calculated based on validity reporting.

Turning to reliability reporting (See Table 3), one sees that studies reporting no reliability data had larger effects than studies with strong reliability reporting, $z(16) = 2.11, p = .04$. However, this result should be interpreted with caution as only one study reported reliability information.

Examining Internal Threats to Validity in Table 3, one notices that studies reporting no internal threats to validity had smaller learning gains than studies reporting two threats, $z(16) = 2.68, p = .01$. No other statistically significant differences were found.

Despite a range of effect size estimates, only studies with zero or one threat to external validity had values statistically greater than 0. In the only significant pairwise comparison (see Figure 5), studies with one external threat to validity had a higher effect size estimate than studies with no threats to validity, $z(8) = 3.29, p = .001$. In contrast to internal threats, there does not appear to be a systematic trend for external threats to validity.
Boxed Hedges’ $g$ values are statistically ($p < .05$) greater than zero, $p < .05$. Diamonds are summary scores, representing Hedges’ $g$ and confidence intervals for two or more outcomes; the final diamond represents all outcomes.

Research question 3: To what extent does assessment level (concept, principle, application) influence cognitive outcomes?

There were no significant differences between assessment levels (see Table 4). However, trends favored Application and Principles-level over Concept-level. With more studies included, the confidence intervals may shrink and significant differences would emerge.

<table>
<thead>
<tr>
<th>Assessment Level</th>
<th>$N_{studies}$</th>
<th>$N_{outcomes}$</th>
<th>$g$</th>
<th>$CI_{Lower}$</th>
<th>$CI_{Upper}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept</td>
<td>3</td>
<td>6</td>
<td>0.46</td>
<td>0.00</td>
<td>0.92</td>
</tr>
<tr>
<td>Principles$^a$</td>
<td>6</td>
<td>8</td>
<td>0.54</td>
<td>0.20</td>
<td>0.88</td>
</tr>
<tr>
<td>Application$^a$</td>
<td>2</td>
<td>3</td>
<td>0.63</td>
<td>0.34</td>
<td>0.93</td>
</tr>
</tbody>
</table>

*Significantly greater than an effect size of 0, $p < .05$. $^aN_{studies}$ does not add up to 7 because some studies employed more than one test that covered two or more assessment levels.

Research question 4: To what extent do the combination of internal threats, external threats, reliability, fading, and scaffolding intervention influence cognitive outcomes?

Regression was used to determine how these variables combine to influence cognitive learning. This involved stepwise regression (forward) selection because including all data with backward elimination would likely result in overfitting due to the small number of observations. A Bonferroni correction was applied for more stringent selection criteria ($t = 1.18$) given the number of variables (Foster & George, 1994). Only variables with statistically significant differences were considered as predictor candidates - reliability reporting, threats to internal validity, threats to external validity, fading, and scaffolding intervention. The latter two were both dummy coded. The final model explains a statistically significant portion of the variance, $R^2 = .30$, $p = .02$ and consists of only a single variable with two levels: fixed fading, or no fading. When scaffolds were not faded, students had higher cognitive outcomes ($t = 2.63$). Scaffolding intervention explained 30% of the variability in cognitive outcomes.
Summary

Overall, scaffolding produced a medium effect \((g = 0.53)\). There were no differences based on study design, assessment level, or intended learning outcome. When examined individually, effect size differences were predicted by reliability reporting, threats to internal validity, external threats to validity, fading, and scaffolding intervention (conceptual or metacognitive). No study reported validity information, and only one study reported reliability information. Studies with zero threats to internal validity had lower effect sizes than studies with two threats. Studies with one threat to external validity had higher effect sizes than studies with zero threats to validity. Studies with no fading had higher effect sizes than studies with fixed fading. Students fared better with conceptual scaffolds than with metacognitive scaffolds. When significant predictors were entered into a regression analysis, fading explained 30% of the variability in cognitive outcomes: students did better in studies with no fading than in studies where scaffolding was faded according to a fixed schedule.

Discussion

Significant differences

One of the most interesting findings was that students fared better when scaffolding did not fade than when it faded on a fixed schedule. The limited number of studies using fixed fading should be noted; including more studies in a larger meta-analysis would help to see if the pattern is consistent. For years, researchers have argued that computer-based scaffolds must fade (Pea, 2004; Puntambekar & Hubscher, 2005). Fading was originally proposed to enable transfer of responsibility from the scaffolding provider to the student (Collins, Brown, & Newman, 1989). According to this reasoning, after fading, students would be able to perform the target task independently. All would agree that promoting transfer of responsibility is important. Some have questioned the theoretical and empirical evidence that fixed fading really leads to transfer of responsibility (Belland, 2011). Without ongoing diagnosis, fading is implemented according to a “best guess” of when scaffolds should be removed by either a designer before a student even uses the scaffold, or by a student as she/he uses the scaffold. Thus, fixed fading may cause scaffolds to be removed (a) too soon, resulting in insufficient support, or (b) too late, resulting in a requirement to use scaffolds when students understand how to perform the target task. In short, if students learn less when using scaffolding with fixed fading, and one cannot be sure that transfer of responsibility happens with fixed fading, then fixed fading becomes less attractive as a scaffolding feature. However, it should be noted that another strategy of fading computer-based scaffolds was not covered in this study – making scaffolds disappear when students indicate they do not need them any more (e.g., Metcalf, 1999).

Results for internal threats to validity and external threats were also interesting. Studies with 2 threats to internal validity had higher effect sizes than studies with 0 threats to internal validity. To promote maximum ecological validity of scaffolding studies among secondary students, studies should be done in middle and high schools. K-12 school administrators and Institutional Review Boards rarely allow individual K-12 students to be randomly assigned to treatments. Thus, research that is conducted in secondary schools will encounter at least one threat to internal validity. Furthermore, by studying scaffolding in laboratory settings where students can be randomly assigned as individuals to treatment conditions, one would lose the essence of scaffolding, as scaffolding is meant to be deployed in authentic, collaborative problem solving or other contexts that require more than just filling out a worksheet or passively listening to or watching instructional materials (Belland, 2014; Wood et al., 1976).

Students did better when there was one external threat to validity than when there were no threats. The two most common threats to external validity in this study were limited description of the scaffolding intervention and experimenter effect. By limited description, we mean that there was enough to determine that it was scaffolding but not enough to replicate the study. A sufficient study procedure description would not change the effect size. Experimenter effect was noted if there was only one instructor. This is common, because early studies of educational interventions often involve a partnership between researchers and one or two teachers (Anderson & Shattuck, 2012). Because teachers do not often design interventions developed in design-based research (Anderson & Shattuck, 2012), fear of artificial inflation of effect sizes through teacher actions is relatively unfounded.

There was a significant difference in cognitive outcomes between conceptual and metacognitive scaffolds. There are two possible explanations: (1) students often do not use metacognitive computer-based scaffolds (Oliver & Hannafin,
(2000), and (2) we examined how scaffolding influences cognitive outcomes. Metacognitive scaffolds may lead to other important outcomes like self-regulated learning ability.

No significant differences

Often it is forgotten that a lack of significant differences can be, in itself, substantial. We did not find any significant differences based on intended learning outcome, assessment level, generic versus context-specific, paired intervention, or study design.

The lack of significant differences based on intended learning outcome provides preliminary evidence that scaffolding works equally well in support of higher-order learning and knowledge integration outcomes. This is encouraging in that students need to gain problem solving ability to be successful in the 21st century workplace (Feller, 2003; Nordgren, 2002), but they also need to recognize that concepts and theories learned in school apply outside of school (Linn, 2000).

Our finding of no significant difference according to assessment level is interesting when compared alongside our previous research in which we found that outcomes of PBL varied based on assessment level (Belland, Walker, Leary, Kuo, & Can, 2010). The inclusion of more studies would help to further elucidate this difference.

That there was no difference in cognitive outcomes between generic scaffolds and context-specific scaffolds is interesting in that it would help scaffold designers make an informed choice about whether to base scaffolding on a generic process or specific content.

It is interesting that there was no difference in learning outcomes according to paired intervention. More studies should be included in a future meta-analysis such that more paired interventions can be included and it can be seen if the trend of no difference in cognitive outcomes holds.

Finally, it is encouraging that there was no difference in cognitive outcomes based on study design because quasi-experimental is the most common quantitative research design in educational research (Gall et al., 2003). This suggests that quasi-experimental designs sufficiently capture the magnitude of scaffolding’s effect.

Remaining issues

While there was a significant difference in study outcomes based on test score reliability, the fact that only one study reported reliability of test scores calls into question the interpretability of this effect. The validity and reliability reporting rate in this study roughly parallels the proportion of PBL studies that reported validity and/or reliability of test scores (Belland, French, & Ertmer, 2009). Unfortunately, reliability and validity reporting is sparse throughout educational research (Belland et al., 2009; Hamdy et al., 2006; Hogan & Agenello, 2004). Readers should recall that reliability and validity reporting should be done not just because the American Educational Research Association (2006) mandates it, but because it has serious consequences for study interpretation. For example, low reliability can lead researchers to under-estimate an effect's magnitude (Hunter & Schmidt, 2004; Loevinger, 1954). If one knows the reliability coefficient, one can estimate the true effect. Low reliability also negatively impacts validity, because one cannot reliably predict what a person with a particular competence level on Trait X will get on an unreliable test that measures Trait X (Messick, 1989). Invalid test scores cannot indicate how much students have of the target construct (Anastasi & Urbina, 1997). Most importantly, improper validity and reliability reporting can lead to improper theory construction based on erroneous empirical results. Improper theory construction has serious consequences, because it can cause school districts and other governmental agencies to spend scarce resources on suboptimal learning tools.

Limitations and suggestions for future research

The final number of included studies is a limitation. We started with 94 studies, but 7 studies were included in the meta-analysis. Eliminated studies (a) did not contain sufficient information to code for an effect size and associated
information, (b) did not have an appropriate control group, or (c) were deemed to not describe a scaffolding intervention. Such elimination required agreement of at least two researchers. Including more studies may have led to different results. Loosening inclusion criteria is not a good choice to increase the number of included articles. Rather, broadening search strategies is. This study was a preliminary meta-analysis intended to optimize our coding and analysis procedures and get a sense of important trends in the scaffolding literature. As such, we limited our selection of studies to those harvested from existing literature reviews. As a next step in our research program, we will conduct a comprehensive search of the primary literature (Cooper, Hedges, & Valentine, 2009). Future research may include search terms describing tools similar to scaffolds, such as mindtools and intelligent tutors. Also, future research should take care to broaden search databases to include ones with greater coverage of the non-USA literature.

Meta-analyses only can include certain quantitative studies. Qualitative studies are common in the scaffolding literature. Empirical studies of a variety of designs and methodologies are all of great value, and all can contribute to an understanding of the impacts of scaffolding. Effect sizes calculated during meta-analyses cannot reflect all pertinent literature on scaffolding. Nonetheless, they can help direct future research and development.

Conclusion and implications

The significance of this paper lies in its affirmation of scaffolding as an effective intervention, and its guidance of future research on scaffolding, as well as of data-driven scaffolding design decisions. First, we found preliminary evidence that scaffolding is effective, producing an average effect size of 0.53. Scaffolding's effect (0.53) is (a) considerably stronger than that of the average instructional intervention designed to promote critical thinking (0.341) (Abrami et al., 2008), yet (b) lower than that found for one-to-one human tutoring (0.79) in a recent meta-analysis (VanLehn, 2011). Still, given the high student-to-teacher ratios in most K-12 schools, and the focus on improving critical thinking abilities of the Common Core State Standards and the Next Generation Science Standards, scaffolding may be a particularly promising intervention (Achieve, 2013; National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010).

Scaffolding produced higher effect sizes when studied in authentic settings (e.g., classroom-based problem-based learning) in which there are more threats to internal and external validity. Thus, educators can have confidence in scaffolding's efficacy even when studies suffered from threats to internal or external validity.

Conceptual scaffolding produced higher effect sizes than metacognitive scaffolds. Scaffolding with no fading produced larger effects than scaffolding with fixed fading. This reinforces the role of teachers in supporting metacognition and transfer of responsibility.

We found preliminary evidence that scaffolding's effectiveness does not depend on whether it (a) supports higher-order outcomes and knowledge integration, (b) is generic or context-specific, (c) is used with different paired interventions, or (d) is assessed at the concept, principles, or application level. These findings may imply that on these characteristics, teachers can select scaffolding that best aligns with learning goals.

Acknowledgments

This work was supported by a grant from Utah State University. Any opinions, findings, and conclusions are those of the authors.

References


A Two-Tier Test-based Approach to Improving Students’ Computer-Programming Skills in a Web-Based Learning Environment

Tzu-Chi Yang1, Gwo-Jen Hwang2, Stephen J. H. Yang1* and Gwo-Haur Hwang3
1Department of Computer Science and Information Engineering, National Central University, No. 300, Jung-Da Road, Chung-li, Taoyuan 320, Taiwan // 2Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, 43, Sec.4, Keelung Rd., Taipei, 106, Taiwan // 3Department of Information Networking and System Administration, Ling Tung University, No. 1, Ling-Tung Rd., Taichung, 400, Taiwan // tcyang.academic@gmail.com // gjhwang.academic@gmail.com // jhyang@csie.ncu.edu.tw // ghhwang@teamail.ltu.edu.tw

*Corresponding author

(Submitted August 15, 2013; Revised February 8, 2014; Accepted June 1, 2014)

ABSTRACT

Computer programming is an important skill for engineering and computer science students. However, teaching and learning programming concepts and skills has been recognized as a great challenge to both teachers and students. Therefore, the development of effective learning strategies and environments for programming courses has become an important issue. To address this issue, this study proposes a two-tier test-based learning system to enhance students’ learning outcomes in computer-programming courses. We conducted an experiment on a college computer-programming course to evaluate the effectiveness of the proposed method. The experimental results show that the proposed method not only improves the students’ attitude toward learning the programming language, but also enhances their programming skills.

Keywords

Computer programming, Computer-assisted learning, Two-tier test, Web-based learning environments

Introduction

The rapid development of information technology has created high demand for skillful programming specialists. Programming skills have therefore become a core competence for engineering and computer science students (Verdú et al., 2012; Hwang, Shadiev, Wang, & Huang, 2012; Fessakis, Gouli, & Mavroudi, 2013). However, learning a computer-programming language involves the comprehension of theoretical background and practice of a range of semantic and syntactic knowledge, coding skills, and algorithmic skills, which are usually complex and difficult for most students to master (Brooks, 1999; Govender, 2009; Katai & Toth, 2010; Wang, Li, Feng, Jiang, & Liu, 2012; Yeh, Chen, Hung, & Hwang, 2010). Researchers have reported that many instructors have encountered difficulties in teaching programming languages. Moreover, most students and teachers agree that learning programming is a challenging task that many students struggle with (Govender & Grayson, 2008; Kordaki, 2010). Therefore, it has become an important and challenging issue to develop effective strategies or tools for teaching computer-programming languages (Emurian, Holden, & Abarbanel, 2008; Hwang, Shadiev, Wang, & Huang, 2012).

Researchers have argued that when learning programming, continuous practice is required to ensure that the knowledge is retained (Chen, Chang, & Wang, 2008; Hwang, Wang, Hwang, Huang, & Huang, 2008). Moreover, actively and periodically scheduled learning is important for students to attain high levels of achievement (Hwang & Wang, 2004). Nevertheless, many computer science students cannot grasp the most fundamental concepts of programming and are thus unable to produce even the most basic programs (Eckerdal, 2009). Researchers have indicated that learning strategy, lack of study, and lack of practice are the causal attributes of success or failure in a computer programming course (Hawi, 2010; Hwang, Wu, Tseng, & Huang, 2011).

Among various learning strategies, providing accurate and meaningful prompts based on individual students’ test results has been recognized by researchers as being an effective approach (Hung, Yang, Fang, Hwang, & Chen, 2014; Hwang, Sung, Hung, Yang, & Huang, 2013; Wu, Hwang, Milrad, Ke, & Huang, 2012). To this end, this study proposes a two-tier test approach with a feedback mechanism. An experiment has been conducted to evaluate the effectiveness of the proposed approach by investigating the following research questions:
• Do the students who learn with the two-tier test-based learning approach show better programming knowledge than those who learn with the conventional technology-enhanced learning approach?
• Do the students who learn with the two-tier test-based learning approach show better programming skills than those who learn with the conventional technology-enhanced learning approach?

**Literature review**

In the past decade, various teaching strategies and learning activities have been applied to computer-programming courses for beginners (Govender & Grayson, 2008). For example, Machanick (2007) proposed the idea of abstraction-first teaching by hiding details until students are ready for them. In addition, Emurian, Holden, and Abarbanel (2008) employed a peer-tutoring approach, and Hwang, Shadiev, Wang, and Huang (2012) proposed a web-based programming-assisted system to provide learning support for programming courses. In the meantime, researchers have indicated that problem-based learning (PBL) could be a promising approach for programming language learning (Kordaki, 2010; Pereira, Zebende, & More, 2010). For example, Gálvez, Guzmán, and Conejo (2009) reported a problem-solving environment to diagnose students’ knowledge levels and to generate feedback and hints to help students understand and overcome their misconceptions in learning programming languages.

According to the literature, providing effective support or guidance by identifying the knowledge levels and learning difficulties of students in learning programming languages could be the key to the improvement in students’ learning performance. Researchers have indicated the important role of assessment in identifying the learning status of individual students in providing effective learning guidance (Hwang, 2003; Hwang, Panjaburee, Triampo, & Shih, 2013; Tseng, Chu, Hwang, & Tsai, 2008). Researchers have further argued that assessment could be used as a learning strategy to improve students’ learning performance (Chu, Hwang, Tsai, & Tseng, 2010; Gálvez, Guzmán, & Conejo, 2009). For example, Hauswirth and Adamoli (2013) proposed a pedagogical approach aligned with this aspect that allows students to learn from their mistakes by answering a series of questions. On the other hand, researchers have also indicated that assessing students’ programming knowledge and skills as well as identifying their learning difficulties remains a challenging issue, implying the need to develop effective strategies or tools to identify students’ status and provide learning support in programming language courses (Wang, Li, Feng, Jiang, & Liu, 2012).

Among the various approaches to programming training, tests are a popular activity adopted by teachers for examining students’ learning status and guiding them to learn (Hwang & Chang, 2011; Gálvez, Guzmán, & Conejo, 2009; Trotman & Handley, 2008; Wang, Li, Feng, & Liu, 2012). Researchers have reported the effectiveness of tests in engaging students to learn and practice, by using multiple-choice items without detailed feedback (Roediger & Karpicke, 2006). The effectiveness of this approach could be attributed to its positive effects on students’ learning retention when engaging them in thinking and practising (Butler, Karpicke, & Roediger, 2007). Gálvez, Guzmán, and Conejo (2009) further suggested that assessment itself can be used as a learning strategy.

Among the existing testing strategies, two-tier tests have been recognized as being an efficient and effective way of investigating students’ prior knowledge or misconceptions, or alternative conceptions by many researchers, especially in science education (Chu & Chang, 2014; Odom & Barrow, 1995; Tsai, 2001). A two-tier test consists of a set of two-level multiple-choice questions. It was first introduced by Treagust (1988), mainly for diagnosing students’ misconceptions or alternative conceptions in science (Tsai & Chou, 2002). The first tier assesses students’ descriptive or factual knowledge of a phenomenon. The second tier probes the students’ reasons for the choice they made in the first tier. The aim of a two-tier test is to explore students’ in-depth explanations of the factual knowledge, which allows teachers or researchers to not only understand students’ possible incorrect ideas, but also to assess the reasoning or in-depth understanding behind these ideas (Chu, Hwang, Tsai, & Tseng, 2010).

In the past decade, researchers have employed the two-tier test approach as an assessment or learning guidance tool. For example, Tsai and Chou (2002) developed a networked two-tier test system in which only one tier of a test item is presented per screen. Such a system facilitates assessment of the existing knowledge of a larger sample of students in a more efficient and relatively straightforward manner. Chou, Chan, and Wu (2007) employed a computerized two-tier test method to assess students’ understanding and alternative conceptions of cyber copyright laws. Researchers have argued that a computerized or web-based two-tier test is not only feasible and efficient, but also provides an easy and familiar interface for students to answer the questions (Chou, Cha, & Wu, 2007). Moreover,
because a two-tier test is in a multiple-choice format, it is much simpler for researchers or educators to interpret students’ responses.

Regarding the cognition levels of programming skills, Brooks (1999) indicated that programming skills can be categorized into four levels (i.e., understanding of the task; finding methods; and coding, testing, and debugging the program). Govender and Grayson (2008) proposed five levels of learning programming (i.e., meeting the requirements, learning syntax, understanding and assimilating, problem solving, and programming in the large). Furthermore, several researchers (Machanick, 2007; Hwang, Wang, Hwang, Huang, & Huang, 2008) have assessed students’ programming skills based on Bloom’s taxonomy; that is, Knowledge, Comprehension, Application, Analysis, Synthesis and Evaluation. Each level of cognitive development is an important dimension for evaluating students’ learning state of programming.

In this study, we used a two-tier test as a strategy for guiding students to learn a programming language. We also developed a web-based assessment and guidance system based on the proposed approach and conducted an experiment to evaluate the learning performance of the students in terms of their programming knowledge (Knowledge and Comprehension levels) and programming skills (Application levels) based on a knowledge test and a skills test.

**Two-tier test-based programming language learning system**

In this study, we propose a two-tier test-based programming language learning system to support web-based learning activities for computer-programming courses. We established a web-based learning environment with client-application-database server architecture. Moreover, we designed the user interface to adapt to the screen sizes of different devices; that is, students are able to interact with the system using personal computers, notebooks, tablets, or smartphones. Individual students can log into the system to review the teaching materials they have learned and to take tests. When students answer a test item, the system provides immediate corresponding feedback, including the correct answer, an explanation of the answer, and the supplementary material for the misconception if the student fails to correctly answer the item.

**Figure 1. System structure of the online two-tier test for programming learning**

**System framework and functions**

We developed a web-based two-tier test system for programming learning, OT3PL, which stands for online two-tier test for programming learning, to assist students in answering questions and justifying their answers with logical arguments. A total of 147 first-tier test items were developed in this study for assessing students’ knowledge and skills of object-oriented programming. Among them, 125 items related to the knowledge level, while 22 items referred to the skills level. Moreover, 452 second-tier items were developed to further confirm the students’
knowledge and ability or to identify their learning difficulties. The proposed system consists of three subsystems: the online test subsystem (OTS), the misconception diagnosis subsystem (MDS), and the learning content interactive subsystem (LCIS). Figure 1 shows the overall framework of the OT3PL.

**Online test subsystem**

After learning a unit or concept, students are asked to finish at least one test before they continue on to the next unit. The online test subsystem aims to generate two kinds of tests: (1) conventional multiple-choice questions selected from the item bank; and (2) a two-tier test, in which every test item has four choices in the first tier, followed by three or four reasons in the second tier for each choice. Table 1 shows an example of a two-tier test item for a programming language. In a conventional multiple-choice item, only the three choices in the first tier (i.e., (A) bar = blue, (B) bar = green, and (C) syntax error) are presented, while an additional nine choices are used to assess the reasons for the students’ choice in a two-tier test item.

**Table 1. Illustrative example of a two-tier test item**

Given the following code:

```java
1. public class Test{
2.    public static void main(String[] args){
3.        String foo = “blue”;
4.        String bar = foo;
5.        foo = “green”;
6.        output.setText (“bar = “+bar);
7.    }
8. }
```

What is the output?

<table>
<thead>
<tr>
<th>First tier (the same in two kinds of test)</th>
<th>Second tier (asked in the two-tier test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) bar = blue</td>
<td>(A1) bar refers to string object “blue” via foo</td>
</tr>
<tr>
<td></td>
<td>(A2) “blue” assigned to bar immediately</td>
</tr>
<tr>
<td></td>
<td>(A3) bar assigned by a copy from foo</td>
</tr>
<tr>
<td>(B) bar = green</td>
<td>(B1) bar assigned to the same memory address as foo</td>
</tr>
<tr>
<td></td>
<td>(B2) bar and foo become the same object since the statement String bar = foo;</td>
</tr>
<tr>
<td></td>
<td>(B3) bar refers to foo</td>
</tr>
<tr>
<td>(C) Syntax error</td>
<td>(C1) The String should be string (in lower-case)</td>
</tr>
<tr>
<td></td>
<td>(C2) Statement “bar = ” + bar is incorrect,</td>
</tr>
<tr>
<td></td>
<td>(C3) In line 5, foo = “green”; causes an error</td>
</tr>
</tbody>
</table>

*Figure 2. Comparison of a conventional online test and a two-tier test*
To prevent students from memorizing the answers, the sequence of the choices and the programming language variables in the test items are changed when generating test sheets. That is, no identical test sheets are generated during the learning process. When making choices in the second tier, students are allowed to view the corresponding first-tier item to identify the reason for making the choice. The scope of the tests is related to the weekly course progress. Students are able to extend the test scope by including more concepts or units that they have learned. Figure 2 shows the system interfaces of the conventional online test and the two-tier test. Figure 3 shows the test results and feedback from the learning system.

Figure 2.

Misconception diagnosis subsystem

In the misconception diagnosis subsystem (MDS), the first-tier item is used to assess the students’ descriptive or factual knowledge of programming problems. The second tier probes the students’ reasons for their choice made in the first tier. Therefore, the choice made in the second-tier item not only confirms whether the students have fully understood a concept, but also helps them to explore their misconceptions if incorrect reasons are selected. Therefore, a decision-tree rule base is used in MDS for diagnosing students’ misconceptions or learning problems. The two-tier test items were developed by two experienced teachers by referring to the past test data in the course. Therefore, misconception identification in this study refers to finding the incorrect concepts, which leads students to make mistakes in programming.

The decision tree for judging the students’ learning problems based on their answers is given in Figure 4. In this example, it is possible for two string objects to have the same value but to be located in different areas of memory in Java. The students may produce a logical error when developing software projects using variables that refer to other variables whose values are changed in later steps if they do not completely understand how to work with objects. For example, in a program, a variable <banana> = “green” and another variable <lemon> = <banana>, the value of <lemon> is therefore “green.” Later, if <banana> is assigned as “yellow,” the students might think that the value of <lemon> is “yellow,” too. In fact, the value of <lemon> remains the same (i.e., “green”). Figure 3 showed that the choice of reason A2 implies that the student is not familiar with the process of string objects, especially value assignment, while if the student takes reason A3 into account, he/she may be confused about how the value of a string object is located in memory.

Figure 3.

Illustrative example of a test result and feedback

The correct answer and explanation to the question.

Immediate feedback provided after the test. The highlight shows that the student has chosen the correct answer but has a misconception about “constructor and method” in JAVA.

The result is “Base.” Two statements in the program may cause this result. The correct reason is “because the default constructor of class alpha is automatically executed after object alpha is created (statement - "new Alpha();"), while class base has no valid default constructor.”

The concepts related to this item.

202
To develop the items for the two-tier test, eight graduate students with more than two years’ programming experience were employed to construct the test items. For example, Table 2 shows four items revised by the teacher in order to identify whether the student completely understands the operator, flow control, and the scope of the variables. The four items in this example are very similar and seem easy to answer. However, even if students are familiar with the use of increasing operators, they will fail to construct flow control statements properly if they do not precisely understand the prefix and postfix of an operator.

<table>
<thead>
<tr>
<th>Table 2. Sample test item</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is the value of i after executing the following code? int i; for(i=1;i&lt;5;i=i+1);</td>
<td>i=5</td>
</tr>
<tr>
<td>What is the value of i after executing the following code? int i=4; for(i=1;i&lt;5;i+=1);</td>
<td>i=5</td>
</tr>
<tr>
<td>What is the value of i after executing the following code? int i; for(i=1;i&lt;5;i=i++);</td>
<td>i=4</td>
</tr>
<tr>
<td>What is the value of i after executing the following code? int i; for(i=1;i&lt;5;i=i++);</td>
<td>unexpected looping</td>
</tr>
</tbody>
</table>

Learning content interactive subsystem

The learning content interactive subsystem (LCIS) presents up-to-date learning contents, including learning material extracted from the textbook and demonstrated programming code provided by the teacher. Students are able to review the correlated contents at any time, such as slides used in the class and the examples given by the teacher. Based on the students’ misconceptions or learning status reported by MDS, LCIS provides relevant supplementary materials to help the students overcome their misconceptions. When students take a test, LCIS presents a list of highlights and provides students with the recommended learning materials. Figure 5 shows an illustrative example of providing personalized learning content to individual students based on test results.
The list of subject units for the student to correct his/her misconceptions.

The student is allowed to select the subject units for further studying.

The list of subject units that the student might not be familiar with.

The learning content related to the use of “array.”

The learning content related to the parameters and return values of a subprogram.

Figure 5. Example of providing personalized comments and learning materials

Experiment design

To evaluate the efficacy of the proposed approach, we conducted an experiment to compare the learning achievements and attitudes of the students who were learning the programming language using the proposed approach. We conducted the experiment on the learning activities of the “Developing Android applications using Java” course of a college in central Taiwan. Details of the experiment are stated below.

Participants

Eighty-eight college students participated in the online programming learning activity. The average age of the students was 20. They were randomly divided into an experimental group and a control group. Eventually, 79 students completed all of the activities, with 40 in the experimental group and 39 in the control group. All of the students were taught by the same teacher.

Treatment

The aim of the course was to train students to develop application programs for mobile devices. The target programming language was Java. The learning objectives included understanding the Java syntax, learning debugging skills, and programming with Java statements to complete specified problems.
Figure 6 presents the experiment design of this study. Both groups of students first received instruction on the basic knowledge of the programming language (Java), then took a pre-test and completed a questionnaire to analyze their knowledge of and attitudes towards the programming language before interacting with the proposed system.

In the first stage of the activity, both groups of students received face-to-face classroom instruction. The teacher provided instruction on the basic syntax and executive environment of Java. The students were then asked to complete a segment of code and perform an exercise based on what they had learned in the classroom.

In the second stage, both groups of students were asked to take a test on the system after each class. In general, they completed at least five exercises and one test each week. In this stage, the students in the experimental group learned with OT3PL. On the other hand, the students in the control group learned with the conventional technology-enhanced learning approach; that is, they were allowed to browse the same learning materials and do the same exercises, but they receive conventional tests and feedback in the web-based learning system.

The learning activity was conducted for four weeks for both groups. After the learning activity, both groups of students took a post-test. They were then asked to fill in the same questionnaire to measure if their attitudes towards programming learning had changed.

Figure 6. Experiment procedure

Measuring tools

Both the pre-test and the post-test were developed by consulting two teachers who had taught the programming language course for more than five years. The pre-test consisted of 33 multiple-choice items, with a total score of 100. The pre-test was designed to evaluate the students’ basic knowledge concerning computer science and programming before the learning activity. The post-test, a detailed achievement test on the Java language, consisted of a programming knowledge test and a programming skills test. The programming knowledge test consisted of 30 multiple-choice items for conceptions, data structure, and coding of Java. The programming skills test consisted of five programming problems. The perfect scores for the two dimensions were 30 and 70, respectively.

The questionnaire for “attitudes toward the programming language course” consisted of seven items. A six-point Likert scheme was used to rate the questionnaire items, where 6 meant strong agreement or positive feedback and 1 represented high disagreement or negative feedback. The Cronbach’s alpha value of the questionnaire was 0.94.
Experimental results

The students’ learning performance was evaluated based on two dimensions, that is, the programming knowledge and programming skills. The former was concerned with the students’ knowledge of programming, such as variables, syntax, and programming concepts. Namely, the programming knowledge dimension focused on the ability of remembering and comprehension of the Java language, and was evaluated by a set of multiple choice items as shown in Table 3.

On the other hand, programming skills refer to the ability of completing software programs based on the specified purposes or function. In other words, to complete a programming task, students require not only programming knowledge and the syntax of the target programming language, but also the ability of identifying problems, organizing codes, and solving problems (Hauswirth & Adamoli, 2013). In this study, the students’ programming skills were evaluated based on the correctness of the programs they developed. For example, they were asked to complete several Android application programs, such as “Please design project T3, such that the screen presents the user’s name in blue when the user touches the screen, but presents the user’s name in red when the user touches the screen for more than two seconds” and “Given the following four pictures, please design project T4, such that users can switch pictures by sliding their finger.” Based on Bloom’s taxonomy, programming knowledge involves recognizing and remembering the usage of Java statements, which develops the Knowledge level and the Comprehension level. Moreover, the programming skills required include solving problems and organizing code statements, relating mainly to the Application level (Anderson & Krathwohl, 2001; Bloom, 1994).

Table 3. Examples of multiple choice items for evaluating programming knowledge

<table>
<thead>
<tr>
<th>Item 10. What is the result of calling the following codes?</th>
</tr>
</thead>
<tbody>
<tr>
<td>int i=1,j=10;</td>
</tr>
<tr>
<td>do{</td>
</tr>
<tr>
<td>if (i&gt;j) {</td>
</tr>
<tr>
<td>break;</td>
</tr>
<tr>
<td>}</td>
</tr>
<tr>
<td>j--;</td>
</tr>
<tr>
<td>}while(++i&lt;5);</td>
</tr>
<tr>
<td>Output.setText(“i=“+i+”and j=“+j);</td>
</tr>
<tr>
<td>A. i = 6 and j = 5</td>
</tr>
<tr>
<td>B. i = 5 and j = 5</td>
</tr>
<tr>
<td>C. i = 6 and j = 4</td>
</tr>
<tr>
<td>D. i = 5 and j = 6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item 21. What is the result of calling the following codes?</th>
</tr>
</thead>
<tbody>
<tr>
<td>String s = “hello” +9 +1 ;</td>
</tr>
<tr>
<td>Output.setText (s) ;</td>
</tr>
<tr>
<td>A. Hello91</td>
</tr>
<tr>
<td>B. Hello10</td>
</tr>
<tr>
<td>C. Compiling error</td>
</tr>
<tr>
<td>D. Throws an exception</td>
</tr>
</tbody>
</table>

| Item28. public class MethodOver {                         |
|     public void setVar (int a , int b , float c ) {}      |
| }                                                         |
| Which of the following is an overloading method of setVar?
| A. private void setVar (int a , float c , int b ) {}     |
| B. protected void setVar (int a , int b , float c ) {}   |
| C. public int setVar (int a , int b , float c ) {return a;} |
| D. protected void setVar (int a , int b , float c ) {return c;} |

Before participating in the learning activity, the students took a pre-test to evaluate their basic knowledge of computer science and programming languages. The means and standard deviations of the pre-test scores were 67.80 and 13.03 for the experimental group, and 66.87 and 11.49 for the control group. A t-test performed on the pre-test scores showed no significant difference between the pre-test results of the two groups, with $t = 0.335$ and $p > .05$; that is, the two groups of students had equivalent knowledge of the computer language prior to the learning activity.

After conducting the two-tier test learning activity, the two groups’ programming knowledge test and programming skills test scores were compared. It was found that the students in the experimental group had significantly better programming skills than the control group, while no significant difference was found between their programming knowledge test scores, as shown in Table 4. According to the aforementioned, the programming knowledge represents the abilities of remembering and comprehending the Java language, while programming skills refers to the ability of organizing programming codes and solving problems. The two-tier test system aimed to help students identify their misconceptions of programming, which could be related to the syntax and semantics of the programming language, or the ability of organizing the programming codes for dealing with a practical problem.
From the experimental results, we found that the two-tier test system benefited the students more in terms of improving their ability of solving problems by organizing the codes in correct ways. In addition, we observed during the learning activity that one third of the students in the experimental group completed all of the tasks in the programming skills test, while most of the students in the control group completed only three tasks.

It should be noted that, in this study, the two-tier test approach was a learning guidance method rather than a test method. The purpose of using the approach was to help the students identify their learning difficulties or problems such that they were able to efficiently cope with the problems via reading supplementary materials or discussing with peers. The effectiveness of such learning guidance or prompts has been reported by several previous studies in science or mathematics courses (Chu et al., 2010; Panjaburee et al., 2013). The interview results from 11 participants provide further evidence for the effectiveness of the two-tier test approach in helping them comprehend the programming knowledge and improve their programming skills. For example, several students have argued that the learning system improved their programming knowledge and skills.

S01: “There are many statements that I did not really understand, because I do not usually use them. However, I have understood the statements since taking the test and following the guidance provided by the learning system.”
S02: “I always study very hard. However, I did not find my weakness in programming and the misconceptions of the programming language until I learned with this system and did the practice.”
S03: “The system refined my knowledge and improved my ability in the programming areas that I did not learn well.”
S08: “It is not easy to learn a programming language by only reading the materials in books and practising without immediate feedback. It is important to do such practice. The test and feedback process is really helpful to me in improving my programming ability.”

Other students also shared similar thoughts that the learning approach not only helped them identify their learning problems, but also improved their programming ability by organizing what they had learned to solve the problems and complete the projects. From the feedback of the students and the experimental results, it is concluded that the two-tier test approach benefited the students by correcting their misconceptions of programming and guiding them to effectively improve their programming knowledge and skills during the learning process.

Such a finding implies that the OT3PL approach can improve the students’ programming skills by enhancing their logical reasoning and by helping them make connections between the programming language (e.g., syntax and types), programming concepts (e.g., variables, data structure, and control flow) and programming procedures (e.g., solving programming problems).

Table 4. The t-test results of the programming knowledge and programming skills tests for the two groups

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean (SD)</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Programming knowledge test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental group</td>
<td>40</td>
<td>20.49 (2.95)</td>
<td>−1.442</td>
</tr>
<tr>
<td>Control group</td>
<td>39</td>
<td>21.48 (3.12)</td>
<td></td>
</tr>
<tr>
<td>Programming skills test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental group</td>
<td>40</td>
<td>55.05 (21.00)</td>
<td>2.546*</td>
</tr>
<tr>
<td>Control group</td>
<td>39</td>
<td>44.49 (15.36)</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05.

Table 5 shows the t-test result of the students’ attitudes toward learning the programming language before and after the learning activity. We found that the learning attitude of the students in the experimental group significantly improved after the learning activity, while the change in the control group students’ attitudes was not significant. This finding conforms to previous studies concerning technology-enhanced learning that effective learning guidance strategies or mechanisms are helpful to students in terms of improving their learning attitudes as well as their learning achievements (Chu, Hwang, Tsai & Tseng, 2010; Hwang & Chang, 2011).

Table 5. The paired t-test result of the students’ attitudes toward programming learning before and after the learning activity

<table>
<thead>
<tr>
<th>Attitudes toward programming learning</th>
<th>Experimental group (N = 40)</th>
<th>Control group (N = 39)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD)</td>
<td>t</td>
<td>p</td>
</tr>
<tr>
<td>Before</td>
<td>4.68 (1.26)</td>
<td>−3.36**</td>
</tr>
<tr>
<td>After</td>
<td>5.33 (1.06)</td>
<td>0.003</td>
</tr>
</tbody>
</table>

*p < .01.
Conclusions

In this paper, we proposed a two-tier test approach for programming learning. Moreover, an online two-tier test system for programming learning with misconception diagnostics and a feedback mechanism, OT3PL, has been implemented based on the proposed approach. OT3PL assists students in assessing their misconceptions of knowledge of programming and in improving their programming skills by judging numbers of statements of code and verifying their reasons. We conducted an experiment on a course on developing Android applications to evaluate the effectiveness of the proposed approach by comparing the learning performance of the students who learned with OT3PL and those who learned with the conventional technology-enhanced learning approach. From the experimental results, we found that, with the help of OT3PL, the experimental group students showed significantly better programming skills than those in the control group.

The finding implies that the two-tier test approach can benefit the students in terms of improving their procedure knowledge rather than their declarative knowledge. Procedural knowledge is the type of knowledge related to how to perform a task, solve problems, or complete a project (Chu, Hwang, Huang, & Wu, 2008; Lewicki, Hill, & Bizot, 1988). In this study, the programming problems in the programming skills test engaged the students in organizing the aims and functions of the problem and applying various programming concepts, syntax, and flow control skills to the development of the program; therefore, the test is relevant to procedural knowledge as well as to the application level of cognition in Bloom’s taxonomy. This finding is reasonable since the two-tier test approach can help students find the learning problems through the interactive questioning and feedback-providing mechanism. On the other hand, the programming knowledge test is related to declarative knowledge, which represents the description of facts or information and is related to the Knowledge and Comprehension levels of cognition. It is also reasonable to find that the two groups of students did not show significant difference in terms of programming knowledge since both approaches (two-tier test-based learning and conventional online test learning) provided the same learning materials to the students to memorize and comprehend.

Although OT3PL benefited the students in this application, there are some limitations to be noted. First, generalization of the findings may be limited to populations of a similar nature, but may not be very applicable for other learner groups with different educational settings or cultural backgrounds. Secondly, the number of students participating is rather small. Third, to provide feedback and judgment, the teachers need to spend time reviewing and refining the test items for evaluation purposes, and the digital learning materials for providing learning supports.

Further research is therefore needed in this area, particularly concerning students of diverse academic degrees and knowledge background. Moreover, it is worth extending this study to the Analysis and Synthesis levels by engaging students in analyzing and improving some existing programs, for example, by reorganizing and reducing code statements for improving the efficiency of the programs. On the other hand, it is also interesting to provide other learning supports, such as collaboration and annotation mechanisms to the learners (Sung & Hwang, 2013; Yang, Yu, & Sun, 2013). In the near future, we will apply this approach to other programming courses to investigate in depth how students look at the system and how they think the system helps them in our future work. Moreover, we plan to investigate the patterns of online behaviors and identify which patterns of behavior possibly result in different learning performance in such a learning activity.

Acknowledgements

This study was supported in part by the National Science Council of the Republic of China under contract numbers NSC 101-2511-S-008-011-MY3, and NSC 101-2511-S-011-005-MY3.

References


The Exploration of Elementary School Teachers’ Internet Self-Efficacy and Information Commitments: A Study in Taiwan

Ying-Tien Wu * and Li-Jen Wang
National Central University, Graduate Institute of Network Learning Technology, No.300, Jhongda Rd., Jhongli, Distict, Taoyuan City 32001, Taiwan // ytwu@cl.ncu.edu.tw // wanglr38@cl.ncu.edu.tw
*Corresponding author

(Submitted October 27, 2013; Revised January 28, 2014; Accepted February 26, 2014)

ABSTRACT
This study aimed to explore teachers’ Internet self-efficacy and information commitments. More importantly, this study also attempted to identify possible factors that affect the teachers’ Internet self-efficacy. The participants were 301 elementary school teachers. In this study, the Internet Self-efficacy Survey (ISS) and the Information Commitments Survey (ICS) were used. Through both exploration factor analyses (EFA) and confirmatory factor analyses (CFA), this study reconfirmed the validity and reliability of the two instruments. This study revealed that the teachers, in general, expressed satisfactory Internet self-efficacy and relatively tended to use advanced information commitments. The correlational analyses revealed that the teachers’ information commitments were significantly correlated with their Internet self-efficacy (p < .05). In order to examine if the teachers’ information commitments were influential factors for their Internet self-efficacy, a series of structural equation modeling (SEM) analyses was conducted. The SEM analyses further confirmed that the teachers’ information commitments were significant factors affecting their Internet self-efficacy. The finding above was then discussed based on Bandura’s self-efficacy theory. Also, some important implications on educational practices and future research were discussed.

Keywords
Internet self-efficacy, Information commitments, Internet-based instruction, In-service teacher

Introduction

With the advancements in information and communication technologies, the integration of technology into education has received more and more attention. In recent years, Internet-based learning has been broadly advocated, and teachers often make extensive use of the Internet before or during instruction (Laxman, 2012; Moore & Chae, 2007; Tsai, 2011). However, as revealed in Childs, Sorensen, and Twidle (2011), teachers often confront some practical but harsh challenges. While they are increasingly taking advantage of the Internet before their instruction, they may not have sufficient confidence to use the Internet when preparing teaching materials and making instructional plans. Also, they may lack confidence in integrating the use of the Internet into their instruction to improve student learning outcomes. In addition, they may have problems in searching for information or resources on the Internet efficiently and making good use of them.

These practical challenges that teachers might face in Internet-based instruction should be highlighted by teacher educators. However, studies addressing the aforementioned issues have not kept pace with teachers’ implementation of Internet-based instruction. In recent years, relevant research has been conducted to explore teachers’ experiences (e.g., Luan et al., 2005), attitudes (e.g., Oral, 2008), views (e.g., Madden et al., 2005), and beliefs (e.g., Savasci-Acikalin, 2009) regarding the Internet or Internet-based instruction. However, relatively less research has addressed teachers’ self-efficacy of exploiting the Internet, efficient use of searching strategies, and appropriate evaluation of online information and resources. As one of the initial attempts, this study was conducted to investigate a group of elementary school teachers’ Internet self-efficacy and information commitments. Also, this study attempted to identify if teachers’ information commitments were possible factors that affected their Internet self-efficacy.

Literature review

Teachers’ Internet self-efficacy

During the last three decades, self-efficacy has received ample attention from educators and researchers, and a great
number of studies related to self-efficacy were conducted (Usher & Pajares, 2008). As conceptualized by Bandura (1994), “Self-efficacy” is an individual’s belief in his/her ability to successfully perform tasks of a particular domain (Bandura, 1994). In general, self-efficacy refers to how confident an individual feels about handling particular tasks, challenges, and contexts (Bandura, 1997). According to Bandura’s (1986) social cognitive theory, an individual’s self-efficacy belief influences his/her choice of activities, how much effort he/she will expend, and how long he/she will sustain effort in dealing with stressful situations. Pajares and Schunk (2001) advocated that, instead of being evaluated in general, research regarding self-efficacy should be assessed at a domain-specific or task-specific level because such measures may have greater validity and predictive relevance. In other words, the exploration of domain-specific or task-specific self-efficacy should be one of the important directions of further self-efficacy research.

For teacher educators, teachers’ self-efficacy is always one of the main concerning issues. In conventional educational contexts, a great amount of research related to teachers’ self-efficacy has been conducted, and these studies have confirmed that teachers’ self-efficacy affects their teaching performance as well as students’ learning outcomes (Labone, 2004). With the implementation of Internet-based instruction, teachers’ Internet self-efficacy has also received more and more attention. Internet self-efficacy refers to an individual’s self-perceived confidence and expectations of using the Internet to complete tasks (Wu & Tsai, 2006). Previous studies have confirmed that individuals with high efficacy expectations of using the Internet will have a greater chance of succeeding in Internet-related tasks (e.g. Tsai & Tsai, 2003). It seems that teachers’ sufficient confidence in using the Internet may be a prerequisite for successful Internet-based instruction. Therefore, an understanding of teachers’ Internet self-efficacy is of much importance for teacher educators and researchers. However, only a few studies have been conducted to address teachers’ Internet self-efficacy recently. For example, Sorensen et al. (2007) revealed that student-teachers’ confidence in the use of the Internet directly influences their application of the Internet in science classes. Kao and Tsai (2009) found that teachers’ Internet self-efficacy could significantly predict their attitudes toward online professional development. Further exploration on teachers’ Internet self-efficacy is crucial.

**Teachers’ information commitments**

In Internet-based instruction, teachers often have to search for and make use of online resources before and during their instruction. Moreover, they often ask students to complete learning tasks in which they have to search for specific information on the Internet and summarize their searching outcomes into reports. Students often encounter disorientation problems (Zhang & Wang, 2010), and lack sufficient skills to evaluate the online information they have found (Child et al., 2011). Therefore, teachers have to help students develop searching skills and make good use of online information by modeling instruction in which they demonstrate how they make use of various searching strategies and how they evaluate the information they have found during online searching processes (Brown, 2001). However, as revealed in Childs et al. (2011), teachers themselves often have problems in searching for information or resources on the Internet efficiently and making good use of them. The situation above is worthy of particular attention from teacher educators and researchers.

To address the aforementioned issues, an exploration of teachers’ “information commitments” would be helpful. Tsai (2004) proposed a theoretical framework for describing learners’ information commitments in Internet-based learning environments, including the three aspects:

- **Standards for accuracy:** The standards that a learner uses to judge the accuracy of online information. The possible orientations are “multiple sources” versus “authority.”
- **Standards for usefulness:** The standards that a learner uses to evaluate the usefulness of online information. The possible orientations are “content” versus “technical.”
- **Searching strategy:** The searching strategies that an individual learner employs to seek online information. The possible orientations are “elaboration” versus “match.”

In recent years, several studies on information commitments have been conducted. However, all of the relevant studies merely focused on exploring learners’ information commitments (e.g., Wu & Tsai, 2005, 2007; Liang & Tsai, 2009). Empirical studies aiming to explore teachers’ information commitments are still not yet available. To address this important issue, this study also investigated teachers’ information commitments.
Investigating possible factors affecting teachers’ Internet self-efficacy

Bandura (1986, 1997) identified four factors affecting the development of one’s self-efficacy: mastery experience, vicarious experience, social persuasion, and physiological state. Among these four factors, mastery experience is the most important one determining an individual’s self-efficacy. How mastery experience could influence an individual learner’s self-efficacy? After completing a learning task, an individual learner may review the results and generate his/her own interpretations. Then he/she creates or revises these interpretations based on his/her judgments of his/her own competence. When he/she believes that his/her efforts have been successful, his/her confidence to accomplish similar or related tasks is raised; on the contrary, when he/she believes that his/her efforts failed to produce the desired effect, his/her confidence to succeed in similar endeavors is diminished.

Researchers have employed a number of studies to examine the applicability of Bandura’s theory about the important factors influencing self-efficacy in conventional educational contexts (Usher & Pajares, 2008). Undoubtedly, Internet self-efficacy could be regarded as a task-specific self-efficacy. Previous research concerned with Internet self-efficacy mostly focused on the exploration of Internet self-efficacy (e.g., Torkzadeh & Van Dyke, 2001), how an individual’s Internet self-efficacy was correlated to or affect his/her Internet-related attitudes, strategies, and behaviors (e.g., Kao & Tsai, 2009; Tsai & Tsai, 2003; 2010), and effects of training on improving Internet self-efficacy (e.g., Torkzadeh & Van Dyke, 2002). The re-examination of the applicability of Bandura’s theory in specific educational contexts or the identification of other possible factors affecting self-efficacy in task-specific contexts (in particular, Internet-based learning) still remain to be investigated in further research. To fill this gap, this study aimed to identify possible factors affecting teachers’ Internet self-efficacy.

Based on Tsai’s (2004) perspectives on information commitments, if teachers utilize advanced searching strategies and employ advanced online information strategies that are commonly used by experts, they are likely to obtain more desired or successful outcomes, which may bring about mastery experience of using the Internet to complete tasks. In other words, teachers’ advanced information commitments may cause their mastery experience regarding Internet-related tasks which may, in turn, strengthen their Internet self-efficacy. On the contrary, if teachers hold less sophisticated information commitments, they may obtain undesired or less successful outcomes in Internet-related tasks, which may diminish their Internet self-efficacy. Therefore, this study considered teachers’ information commitments as a possible factor affecting their Internet self-efficacy, and the perspectives above were further examined in the current study.

Research questions

The research questions of this current study are:
- What are the elementary school teachers’ Internet self-efficacy and their information commitments?
- What are the relationships (if any) between elementary school teachers’ Internet self-efficacy and their information commitments?
- If the relationships mentioned in research question 2 do exist, do elementary school teachers’ information commitments have significant effects on their Internet self-efficacy?

Method

Sample

In this study, the participants were 301 voluntary elementary school teachers in Taiwan. Among these teachers, 92 were male (30.6%) while the remaining 209 were female (69.4%). These teachers had differing degrees of Internet experience: 169 reported less than 12 online hours per week, 79 reported 13-24 hours, and 53 reported more than 24 hours per week.

Instruments

In this study, to assess the teachers’ Internet self-efficacy and their information commitments, two instruments, the
Internet Self-efficacy Survey (ISS) and the Information Commitments Survey (ICS), were used. The ISS developed in Kao, Wu and Tsai (2011) consists of two scales (a total of 14 items). It is presented with bipolar strongly confident/strongly unconfident statements in a seven-point Likert scale. The details of the two scales are as follows:

- **Basic self-efficacy scale (7 items):** measuring teachers’ perceived confidence of using the Internet at a basic level, such as using Internet-related tools. The higher the score a teacher obtains, the better his/her basic self-efficacy regarding using the Internet. A sample item is “I feel confident of printing out the content of a Website.”
- **Advanced self-efficacy scale (7 items):** assessing teachers’ perceived confidence and self-expectations of Internet-based interaction or advanced usages of the Internet, such as searching for online information with keywords or using Internet-based communication tools. The higher the score a teacher has, the more confidence in advanced usage of the Internet he/she has. A sample item is “I feel confident of offering information or answering questions on the Internet.”

In this study, the ICS was used to investigate the elementary school teachers’ information commitments. Based on Tsai’s (2004) theoretical framework regarding information commitments, the ICS was originally developed in Wu and Tsai (2005, 2007). The ICS consists of six scales (a total of 24 items), and is presented with bipolar strongly agree/strongly disagree statements in a seven-point Likert scale. The six scales are:

- **Multiple sources scale (3 items):** measuring to what extent a teacher will evaluate the correctness of unknown online information he/she has found by relating to other websites, his/her prior knowledge, peers, or other printed materials. A sample item is “When I view some unknown information on the Internet, I will explore relevant content from books (or printed materials), and then evaluate whether the information is correct.”
- **Authority scale (4 items):** assessing the extent to which a teacher will examine the accuracy of unknown information in Internet-based learning environments by the “authority” of the websites or other sources. A sample item is “When I view some unknown information on the Internet, I will believe in its accuracy if the information appears in some websites recommended by experts.”
- **Content scale (5 items):** measuring to what extent a teacher will assess the usefulness of the online information he/she has found by the relevancy of its content. A sample item is “When I view of search for the information on the Internet, the information is useful for me if it is highly related to my intended searching content.”
- **Technical scale (4 items):** assessing the extent to which a teacher will judge the usefulness of the online information he/she has found by the ease of retrieval, the ease of searching or the ease of obtaining information. That is, their standard for evaluating the usefulness of online information is more related to some technical issues. A sample item is “When I view of search for the information on the Internet, the information is useful for me if the interface is well designed.”
- **Elaboration scale (5 items):** measuring to what extent a teacher will have purposeful (metacognitive) thinking or integrate online information or resources from several websites to find the best fit that fulfills his/her purpose. A sample item is “When I need to search for information on the Internet, I am used to summarizing a variety of information.”
- **Match scale (3 items):** investigating the extent to which a teacher will be eager to find only a few websites that contain the most fruitful and relevant information when searching for online information. A sample item is “When I need to search for information on the Internet, I will not search for other sites once I find the first relevant website.”

**Data analyses**

Both exploratory factor analyses (EFA) and confirmatory factor analyses (CFA) were conducted in this study to clarify and confirm the structure of teachers’ Internet self-efficacy and the structure of their information commitments. When conducting the EFA in this study, principle component analysis was used as the extraction method, and varimax with Kaiser normalization was used as the rotation method. An item was retained only when it loaded greater than 0.5 on the relevant scale and smaller than 0.5 on other unrelated scales. Moreover, with structural equation modeling (SEM) analyses, two series of CFA were also conducted to confirm the reliability and validity of the ICS and the ISS respectively.

Besides, a series of SEM analyses was also conducted in this study to examine if the elementary school teachers’ information commitments had statistically significant effects on their Internet self-efficacy. Multiple fit indices were used to evaluate the fitness of the structural model obtained by the SEM analyses. These indices and the recommended values of these indices are shown in Table 5.
Results

Instrument validation

According to Table 1, the teachers’ responses to the ISS can be grouped into the two factors, “Basic Internet self-efficacy” (7 items) and “Advanced Internet self-efficacy” (7 items). The reliability (alpha) coefficients for the two scales were 0.93 and 0.94 respectively, and the overall alpha of the ISS was 0.93. The two factors accounted for a total variance of 73.43%. Therefore, the ISS used in this study was sufficiently reliable for assessing elementary school teachers’ Internet self-efficacy.

Moreover, by way of a series of CFA, the fitness of items for each ISS scale was examined. Table 1 revealed that all t-values of the 14 items on the two ISS scales were significant, indicating that the ICS used in this study had high convergent validity ($p < .05$). As shown in Table 1, the factor loadings of all the items were higher than 0.5, revealing satisfactory factor loading in its measuring scale. In addition, the results in Table 1 show that the composite reliability coefficients of the two scales were 0.93 and 0.94, respectively. They were all larger than 0.6, revealing high composite reliability. In sum, the ISS used in this study has high validity and sufficient reliability for assessing elementary school teachers’ Internet self-efficacy.

Table 1. A summary of the EFA and CFA results for the ISS ($n = 301$)

<table>
<thead>
<tr>
<th>Scale</th>
<th>Item</th>
<th>Mean</th>
<th>SD</th>
<th>factor loading</th>
<th>Eigen-value</th>
<th>variance explained</th>
<th>Cronbach α</th>
<th>factor loading</th>
<th>t-value</th>
<th>composite reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic Internet self-efficacy</td>
<td>BS1</td>
<td>6.64</td>
<td>0.73</td>
<td>0.79</td>
<td>8.16</td>
<td>58.23%</td>
<td>0.93</td>
<td>0.78</td>
<td>15.82*</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>BS2</td>
<td>6.66</td>
<td>0.59</td>
<td>0.76</td>
<td></td>
<td></td>
<td></td>
<td>0.73</td>
<td>14.41*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BS3</td>
<td>6.55</td>
<td>0.91</td>
<td>0.86</td>
<td></td>
<td></td>
<td></td>
<td>0.91</td>
<td>20.20*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BS4</td>
<td>6.61</td>
<td>0.80</td>
<td>0.88</td>
<td></td>
<td></td>
<td></td>
<td>0.91</td>
<td>20.49*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BS5</td>
<td>6.48</td>
<td>0.86</td>
<td>0.77</td>
<td></td>
<td></td>
<td></td>
<td>0.78</td>
<td>15.81*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BS6</td>
<td>6.55</td>
<td>0.74</td>
<td>0.79</td>
<td></td>
<td></td>
<td></td>
<td>0.83</td>
<td>17.54*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>BS7</td>
<td>6.59</td>
<td>0.74</td>
<td>0.80</td>
<td></td>
<td></td>
<td></td>
<td>0.79</td>
<td>16.14*</td>
<td></td>
</tr>
<tr>
<td>Advanced Internet self-efficacy</td>
<td>AS1</td>
<td>5.78</td>
<td>1.60</td>
<td>0.81</td>
<td>2.13</td>
<td>15.18%</td>
<td>0.94</td>
<td>0.90</td>
<td>19.81*</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>AS2</td>
<td>5.88</td>
<td>1.52</td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
<td>0.87</td>
<td>18.79*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AS3</td>
<td>5.88</td>
<td>1.48</td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
<td>0.86</td>
<td>18.69*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AS4</td>
<td>5.70</td>
<td>1.70</td>
<td>0.89</td>
<td></td>
<td></td>
<td></td>
<td>0.92</td>
<td>20.88*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AS5</td>
<td>5.88</td>
<td>1.54</td>
<td>0.76</td>
<td></td>
<td></td>
<td></td>
<td>0.73</td>
<td>14.64*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AS6</td>
<td>5.44</td>
<td>1.75</td>
<td>0.85</td>
<td></td>
<td></td>
<td></td>
<td>0.77</td>
<td>15.70*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AS7</td>
<td>5.15</td>
<td>1.86</td>
<td>0.83</td>
<td></td>
<td></td>
<td></td>
<td>0.74</td>
<td>14.92*</td>
<td></td>
</tr>
</tbody>
</table>

$\text{Overall } \alpha = 0.93 \text{ and total variance explained } = 73.43\%.$

$p < .05.$

According to the EFA results shown in Table 2, the teachers’ responses on the items of the ICS can be grouped into the six factors: “multiple sources,” “authority,” “content,” “technical,” “elaboration,” and “match.” The reliability (alpha) coefficients for these scales respectively were 0.82, 0.89, 0.93, 0.82, 0.91 and 0.95, and the overall alpha value of the ICS was 0.90. Exactly as in the ICS developed by Wu and Tsai (2007), there are 3, 4, 5, 4, 5, and 3 items in each of the ICS scales. Besides, the six factors of ICS accounted for 78.23% of the variance. Therefore, the ICS was deemed to be sufficiently reliable for assessing elementary school teachers’ information commitments.

Table 2 also reveals the results of CFA on the ICS. According to Table 2, all t-values of the 24 items in ICS showed significance at the 0.05 level, suggesting that the ICS revealed high convergent validity. The factor loadings of the 24 items were greater than 0.5, indicating high factor loading in its measuring scale. Also, the reliability of ICS was evaluated again by assessing the composite reliability coefficient of each scale. As shown in Table 2, the composite reliability coefficients of the six scales were 0.83, 0.90, 0.93, 0.82, 0.91 and 0.95, and were all larger than 0.6. In sum, the ICS used in this study has high validity and sufficient reliability for assessing elementary school teachers’ information commitments.
Table 2. A summary of the EFA and CFA results for the ICS (n = 301)

<table>
<thead>
<tr>
<th>Scale</th>
<th>Item</th>
<th>Mean</th>
<th>SD</th>
<th>EFA</th>
<th>CFA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Exploratory factor analyses</td>
<td>Confirmatory factor analyses</td>
</tr>
<tr>
<td></td>
<td>Item</td>
<td>factor loading</td>
<td>Eigenvalue</td>
<td>variance explained</td>
<td>Cronbach α</td>
</tr>
<tr>
<td>Multiple Sources</td>
<td>Multiple 1</td>
<td>5.78</td>
<td>0.78</td>
<td>0.77</td>
<td>1.18</td>
</tr>
<tr>
<td></td>
<td>Multiple 2</td>
<td>5.76</td>
<td>0.88</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Multiple 3</td>
<td>5.81</td>
<td>0.94</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>Authority</td>
<td>Authority 1</td>
<td>5.26</td>
<td>1.05</td>
<td>0.78</td>
<td>2.44</td>
</tr>
<tr>
<td></td>
<td>Authority 2</td>
<td>5.26</td>
<td>1.12</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Authority 3</td>
<td>5.25</td>
<td>0.97</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Authority 4</td>
<td>5.43</td>
<td>0.91</td>
<td>0.85</td>
<td></td>
</tr>
<tr>
<td>Content</td>
<td>Content 1</td>
<td>6.04</td>
<td>0.66</td>
<td>0.82</td>
<td>8.31</td>
</tr>
<tr>
<td></td>
<td>Content 2</td>
<td>5.91</td>
<td>0.71</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Content 3</td>
<td>5.94</td>
<td>0.68</td>
<td>0.85</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Content 4</td>
<td>6.08</td>
<td>0.81</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Content 5</td>
<td>6.03</td>
<td>0.61</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>Technical</td>
<td>Technical 1</td>
<td>5.13</td>
<td>1.08</td>
<td>0.63</td>
<td>1.27</td>
</tr>
<tr>
<td></td>
<td>Technical 2</td>
<td>5.43</td>
<td>1.28</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Technical 3</td>
<td>5.39</td>
<td>1.19</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Technical 4</td>
<td>4.32</td>
<td>1.51</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td>Elaboration</td>
<td>Elaboration 1</td>
<td>5.81</td>
<td>0.80</td>
<td>0.79</td>
<td>3.61</td>
</tr>
<tr>
<td></td>
<td>Elaboration 2</td>
<td>5.90</td>
<td>0.70</td>
<td>0.78</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Elaboration 3</td>
<td>5.88</td>
<td>0.81</td>
<td>0.81</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Elaboration 4</td>
<td>5.68</td>
<td>0.81</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Elaboration 5</td>
<td>5.79</td>
<td>0.80</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>Match</td>
<td>Match 1</td>
<td>4.13</td>
<td>1.44</td>
<td>0.92</td>
<td>1.96</td>
</tr>
<tr>
<td></td>
<td>Match 2</td>
<td>3.63</td>
<td>1.55</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Match 3</td>
<td>3.76</td>
<td>1.42</td>
<td>0.91</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05.

Overall α = 0.90 and total variance explained = 78.23%.

Teachers' scores on the ISS and ICS

Table 3 shows the teachers’ mean scores and the standard deviations of the two scales of the ISS. The elementary school teachers in this study, on average, scored higher than five points on both of the two ISS scales. They scored higher on the “basic self-efficacy” scale (M = 6.58, SD = 0.65) than on the “advanced self-efficacy” scale (M = 5.67, SD = 1.40), implying that the teachers in this study displayed high confidence and expectations of using the Internet for both basic and advanced purposes, but they had relatively higher confidence and expectations of using the Internet for basic purposes than for advanced purposes.

According to Table 3, on the six scales of the ICS, the teachers scored highest on “content” (M = 6.00, SD = 0.62), followed by “elaboration” (M = 5.81, SD = 0.67), “multiple sources” (M = 5.78, SD= 0.74), “authority” (M = 5.30, SD = 0.88), “technical” (M = 5.07, SD = 1.02), and “match” (M = 3.84, SD = 1.40). It seems that the teachers in this study had relatively greater tendencies to judge the accuracy and usefulness of online information or resources with the advanced information commitments (i.e., “content” and “multiple sources”).

As for searching strategies, the teachers expressed great orientation in using the “elaboration” searching strategy (M = 5.81, SD = 0.67), indicating that they were oriented towards making use of sophisticated searching strategies. Besides, it should be noticed that the mean score of the teachers’ responses on the “match” scale (M = 3.84) is lower...
than 4 (the middle value in a seven-point Likert scale). That is, these teachers did not agree that they often employed the “match” searching strategy, which was categorized as one of the less sophisticated information commitments in previous research. In sum, the teachers in this study tended to use sophisticated searching strategies (such as “elaboration”), rather than less advanced ones (such as “match”).

**Table 3. Teachers’ scores on the ISS and ICS (n = 301)**

<table>
<thead>
<tr>
<th>ISS scale</th>
<th># of items</th>
<th>Item mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic self-efficacy</td>
<td>7</td>
<td>6.58</td>
<td>0.65</td>
</tr>
<tr>
<td>Advanced self-efficacy</td>
<td>7</td>
<td>5.67</td>
<td>1.40</td>
</tr>
</tbody>
</table>

**ICS scale            | # of items | Item mean | SD  |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple sources</td>
<td>3</td>
<td>5.78</td>
<td>0.74</td>
</tr>
<tr>
<td>Authority</td>
<td>4</td>
<td>5.30</td>
<td>0.88</td>
</tr>
<tr>
<td>Content</td>
<td>5</td>
<td>6.00</td>
<td>0.62</td>
</tr>
<tr>
<td>Technical</td>
<td>4</td>
<td>5.07</td>
<td>1.02</td>
</tr>
<tr>
<td>Elaboration</td>
<td>5</td>
<td>5.81</td>
<td>0.74</td>
</tr>
<tr>
<td>Match</td>
<td>3</td>
<td>3.84</td>
<td>1.40</td>
</tr>
</tbody>
</table>

**The correlation between teachers’ Internet self-efficacy and their information commitments**

According to Table 4, the teachers’ responses on the “Multiple sources,” the “Content,” and the “Elaboration” scales of the ICS were all significantly positively correlated with their responses on the “General self-efficacy” scale of ISS (p < .05). Also, their responses on the three scales of ICS (i.e., “Multiple sources,” “Content,” and “Elaboration” scales) were all significantly positively correlated with their responses on the “Advanced self-efficacy” scale of ISS (p < .05). According to Tsai (2004), the three information commitments, “multiple sources,” “content,” and “elaboration,” are commonly expressed by experts, and are categorized as advanced information commitments, while the other three are usually expressed by novices, and are viewed as less sophisticated. The results above indicate that the more the teachers were oriented to evaluate the correctness and the usefulness of online information by using the evaluative standards that the experts commonly use (i.e., evaluate the correctness of online information from multiple sources and judge the usefulness of online information by the relevancy of its content), the greater Internet self-efficacy they would have. Also, the more they were oriented to use the “elaboration” searching strategies that often utilized by experts, the greater Internet self-efficacy they would have. However, Table 4 also reveals that the teachers’ responses on the “Match” scale of ICS were significantly negatively correlated with their responses on both of the two ISS scales (p < .05), indicating that the more the teachers were prone to find only a few websites containing the most fruitful and relevant information, the lower Internet self-efficacy they would express.

**Table 4. The correlation between teachers’ Internet self-efficacy and their information commitments (n = 301)**

<table>
<thead>
<tr>
<th></th>
<th>Multiple sources</th>
<th>Authority</th>
<th>Content</th>
<th>Technical</th>
<th>Elaboration</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic self-efficacy</td>
<td>0.32**</td>
<td>0.06</td>
<td>0.26*</td>
<td>-0.08</td>
<td>0.37**</td>
<td>-0.27**</td>
</tr>
<tr>
<td>Advanced self-efficacy</td>
<td>0.38**</td>
<td>0.08</td>
<td>0.15*</td>
<td>-0.02</td>
<td>0.37**</td>
<td>-0.13**</td>
</tr>
</tbody>
</table>

p < .05. **p < .01.

In sum, the findings above seem to suggest that teachers’ information commitments are correlated with their Internet self-efficacy. In particular, the more teachers tend to hold advanced information commitments that are commonly expressed by experts, the higher Internet self-efficacy they may have, while the more they are oriented to have less sophisticated information commitments that are often used by novices, the lower Internet self-efficacy they may express.

**The causal relations between teachers’ information commitments and their Internet self-efficacy**

To specify that if the teachers’ information commitments were influential factors for their Internet self-efficacy, a series of SEM analyses was further conducted. A structural model, as shown in Figure 1, was obtained from using the six scales of ICS as the predictor variables and the two scales of ISS as the outcome variables for the SEM analyses.
The fit measures for this model are summarized in Table 5. The fit index values reveal that there is a high satisfactory fit between the structural model and the sample data of this study.

Table 5. Fit measures for the structural model

<table>
<thead>
<tr>
<th>Fit index</th>
<th>Structural model</th>
<th>Recommended value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x^2$/ DF</td>
<td>4.50</td>
<td>≤ 5</td>
</tr>
<tr>
<td>Root Mean Square Error of Approximation (RMSEA)</td>
<td>0.11</td>
<td>≤ 0.08</td>
</tr>
<tr>
<td>Root Mean Square Residual (RMR)</td>
<td>0.06</td>
<td>≤ 0.10</td>
</tr>
<tr>
<td>Standard RMR (SRMR)</td>
<td>0.06</td>
<td>≥ 0.05</td>
</tr>
<tr>
<td>Goodness of Fit Index (GFI)</td>
<td>0.94</td>
<td>≥ 0.90</td>
</tr>
<tr>
<td>Adjusted Goodness of Fit Index (AGFI)</td>
<td>0.87</td>
<td>≥ 0.80</td>
</tr>
<tr>
<td>Normed Fit Index (NFI)</td>
<td>0.90</td>
<td>≥ 0.90</td>
</tr>
<tr>
<td>Comparative Fit Index (CFI)</td>
<td>0.92</td>
<td>≥ 0.90</td>
</tr>
<tr>
<td>Incremental Fit Index (IFI)</td>
<td>0.92</td>
<td>≥ 0.90</td>
</tr>
</tbody>
</table>

In Figure 1, the statistically significant relations between teachers’ Internet self-efficacy and information commitments are shown as solid lines. According to Figure 1, both “Multiple sources” and “Elaboration” had significantly positive effects on “Basic Internet self-efficacy” ($p < .05$), as well as on “Advanced Internet self-efficacy” ($p < .05$). However, Figure 1 also shows that “Match” has a significant negative effect on both “Basic Internet self-efficacy” and “Advanced Internet self-efficacy,” indicating that the teachers’ advanced evaluative standards for the accuracy of online information and resources (i.e., “multiple sources”) and their use of sophisticated information searching strategies did improve their Internet self-efficacy; on the other hand, their usage of a less advanced information searching strategy (i.e., “Match”) might diminish their Internet self-efficacy. That is, generally speaking, the teachers’ information commitments did indeed have effects on their Internet self-efficacy. If teachers make use of advanced information commitments that are commonly expressed by experts, their Internet self-efficacy...
may be strengthened; on the contrary, if teachers utilize less sophisticated information commitments (in particular, use less advanced online searching strategies) that are often held by novices, their Internet self-efficacy may be diminished.

**Discussion**

This study is one of the initial attempts addressing teachers’ Internet self-efficacy and information commitments. In this study, a total of 301 elementary school teachers’ Internet self-efficacy and their information commitments were investigated. More importantly, the current study also attempted to identify possible factors that affect the teachers’ Internet self-efficacy. The major findings of this study are discussed below.

**Improving teachers’ Internet self-efficacy**

For teacher educators, teachers’ Internet self-efficacy is crucial and should be highlighted. We can look at this issue at two levels. On one hand, as already discussed in the Introduction section, teachers’ sufficient confidence in using the Internet could be a prerequisite for successful Internet-based instruction; on the other hand, just like learners, teachers may have to utilize the Internet to enhance their profession of teaching when attending online professional development programs. In recent years, online professional development has been increasingly implemented (Dede, Ketelhut, Whitehouse, Breit, & McCloskey, 2009). When teachers undertake online professional development, they could advance their profession of teaching with the Internet, just like learners behave in Internet-based learning environment. As learners’ Internet self-efficacy plays a significant role in their Internet-based learning outcomes, teachers’ satisfactory Internet self-efficacy could be crucial for their online professional development outcomes.

In this study, compared with their basic Internet self-efficacy, these elementary school teachers showed relatively lower confidence and expectations of using the Internet for advanced purposes, such as online searching with keywords and online communication. As teachers’ advanced usage of the Internet, such as online interaction or communication, how to improve their advanced Internet self-efficacy should be highlighted. However, still no empirical research on this particular issue is currently available. Therefore, further research should be conducted to address this important issue. The findings in Torkzadeh and Van Dyke (2002) and this study may provide some possible directions. In Torkzadeh and Van Dyke (2002), it was reported that the training significantly improved university students’ Internet self-efficacy. However, the content of the training program was not clearly reported. In this study, it was revealed that the more teachers were oriented to evaluate the correctness and the usefulness of online information by using the evaluative standards, the greater Internet self-efficacy they would have. Besides, the more they were oriented to use the “elaboration” searching strategies that often utilized by experts, the greater Internet self-efficacy they would have. Based on Torkzadeh and Van Dyke (2002) and the findings above, it seems that training programs for teachers may be helpful for improving their advanced Internet self-efficacy. However, the training programs should focus on how to improve teachers’ use of both advanced online searching strategies and sophisticated online information evaluation. Typically, when designing training programs for improving teachers’ Internet self-efficacy, teacher educators should also pay particular attention to how to help them develop advanced evaluative standards for online information and increase their use of sophisticated online searching strategies.

**Information commitments of Internet users with various specialties or majors**

One may be interested in if teachers’ information commitments are different from other Internet users with various specialties or majors. However, till now, this study is the first attempt to explore teachers’ information commitments, and none study has been conducted to investigate other adult with various specialties or majors. But, the findings in Wu and Tsai (2007) and Liang and Tsai (2009) may provide some possible clues. Both these two studies explored university students’ information commitments; however, the majors of the participants in these two studies were quite different and they were found to hold quite different information commitments. In Wu and Tsai (2007), the participants, who mainly majored in science or engineering, expressed a greater tendency to evaluate the usefulness of online information by using advanced information commitments (i.e., “content”). Also, they were more oriented to use sophisticated information searching strategies (i.e., “elaboration”), rather than less advanced information searching strategy (i.e., “match”). However, in Liang and Tsai (2009), it was found that the medical students focused
more on some technical issues rather than on real content when evaluating the usefulness of online information. They also tended to use the “match” rather than the “elaboration” online searching strategy. It seems that learners’ specialties or majors may be a significant mediated variable for their information commitments. Similarly, teachers’ information commitments may be different from other adults with various specialties or majors. The perspective above is suggested to be confirmed in future research.

Nowadays, except general information, information related to specialized knowledge is more and more readily available for users on the Internet (e.g., Wikipedia, discussion forums and online communities, provide us with bound resources of specialized knowledge). Therefore, except searching for and making use of general information, an individual with a specialty or a profession may also have to search and evaluate information related to his/her profession. It may be interesting to investigate that if an individual holds different information commitments when searching and evaluating general information or information related to his/her specialty on the Internet. Therefore, further studies can be conducted to compare teachers’ information commitments for different kinds of information (i.e., general information and information related to their specialties).

**Information commitments as factors affecting teachers’ Internet self-efficacy**

Bandura (1986, 1997) identified four factors affecting the development of one’s self-efficacy: mastery experience, vicarious experience, social persuasion, and physiological state. Among these four factors, mastery experience is the most important one determining an individual’s self-efficacy. Through the SEM analyses, this study revealed that the teachers’ information commitments had statistically significant effects on their Internet self-efficacy. Particularly, their advanced evaluative standard for the accuracy of online information and resources (i.e., “multiple sources”) and their use of sophisticated information searching strategies (i.e., “elaboration”) improved their Internet self-efficacy, while their use of less advanced information searching strategies (i.e., “match”) diminished their Internet self-efficacy. The findings above, which have identified information commitments as significant factors affecting Internet self-efficacy, could contribute to extend Bandura’s (1986, 1997) self-efficacy theory about the factors affecting self-efficacy in Internet-related contexts.

The SEM analyses conducted in this study solely focused on specifying the causal relations of teachers’ information commitments to their Internet self-efficacy. It is suggested that future research should be conducted to identify other possible factors affecting Internet self-efficacy. If more possible factors are identified, then further research could be conducted to find out stronger factors among them. This will help to examine the applicability of Bandura’s theory in Internet-related contexts, extend his theory, and provide us with deeper insights into Internet self-efficacy.

**Possible reciprocal causation between teachers’ Internet self-efficacy and their online information searching strategies**

In Tsai and Tsai (2003), the influence of university students’ Internet self-efficacy on their use of online searching strategies was explored, and it was revealed that students with high Internet self-efficacy tended to use advanced searching strategies. According to the structure model in Figure 1, the teachers’ usage of advanced information searching strategies had significant effects on their Internet self-efficacy, while their use of less advanced information searching strategies caused of their lower Internet self-efficacy. In sum, the teachers’ use of information searching strategies affected their Internet self-efficacy. According to the current study and Tsai and Tsai (2003), it is likely that an individual’s Internet self-efficacy and his/her use of searching strategies may mutually reinforce one another. In other words, individuals’ Internet self-efficacy and their use of searching strategies may have reciprocal causation. To confirm this perspective, further research is also suggested.

**Conclusions**

To conclude, the elementary school teachers in this study, in general, expressed satisfactory Internet self-efficacy and advanced information commitments. Moreover, the teachers’ information commitments were identified as significant factors affecting their Internet self-efficacy. The findings derived from this study could be helpful in extending Banduras’ theory about the factors affecting self-efficacy in Internet-related contexts. Moreover, this study also
provides some implications in terms of educational practices and future research. For teacher education or teacher professional development, when designing training programs for improving teachers’ Internet self-efficacy, teacher educators should pay particular attention to how to develop teachers’ advanced evaluative standards for online information and increase their use of sophisticated online searching strategies. For instance, advanced evaluative standards for online information and sophisticated online searching strategies should be explicitly introduced to teachers. Also, in these training programs, teachers should be guided to practice making use of their metacognition in Internet-related tasks. Regarding future research, this study also suggests some possible directions. For example, further studies can be conducted to compare teachers’ information commitments for different kinds of information (i.e., general information and information related to their specialties). In addition, identifying other possible factors affecting Internet self-efficacy and investigating stronger factors among these influential factors can be also conducted.

Limitations

For educational researchers, causal relations of constructs are generally derived from the use of experimental research designs and the conduct of carefully manipulated experiments. For years, SEM analyses have been widely used to examine the causality of constructs. With SEM analyses, a structural model specifying the causal relations of the constructs to one another could be obtained (Anderson & Gerbing, 1988). However, some researchers may have concerns about using SEM results to indicate causality, and the adoption of SEM analyses for specifying the causal relations of constructs is controversial (Pearl, 2012). By using SEM analyses, this study initially confirmed the possible effects of teachers’ information commitments on their Internet self-efficacy. It should be acknowledged that due to this study was not conducted with an experimental research design, it only revealed the statistical causal relationship between information commitments and Internet self-efficacy. As a result, the findings derived from SEM analyses in this study should be carefully interpreted.

Acknowledgements

Funding of this research work was supported by the National Science Council, Taiwan, under grant number NSC 101-2628-S-008-001-MY3. The authors would like to express their gratitude to Prof. Chin-Chung Tsai for his helpful comments in the further development of this paper.

References


Social Presence and Interaction in Learning Environments: The Effect on Student Success

Ines Kožuh1*, Zoran Jeremić2, Andrej Sarjaš1, Julija Lapuh Bele3, Vladan Devedžić4 and Matjaž Debevc1

1University of Maribor, Faculty of Electrical Engineering and Computer Science, Smetanova 17, 2000 Maribor, Slovenia // 2Ryerson University, Laboratory for System, Software and Semantics (LS3), Toronto, Ontario Canada // 3B2, Vocational College, Tržaška 42, 1000 Ljubljana, Slovenia // 4University of Belgrade, FON - School of Business Administration, POB 52, Jove Ilica 154, 11000 Belgrade, Serbia // ines.kozuh@um.si // zoran.jeremic@gmail.com // andrej.sarjas@um.si // julija.bele@b2.eu // devedzic@fon.rs // matjaz.debevc@um.si

*Corresponding author

(Submitted October 2, 2013; Revised January 6, 2014; Accepted February 21, 2014)

ABSTRACT

With the increased use of social media there is a growing interest in using social interaction and social presence in education. Despite this phenomenon, no appropriate methodology was found on effective integrating of both concepts into online learning. In this study, we propose integrating two different kinds of learning tools to provide social interaction and social presence in Personal Learning Environments. We have evaluated the proposed concept in a classroom setting, using a specific social interaction tool and a specific social presence tool. The findings revealed that although the use of the social interaction tool was positively associated with students’ academic success, the perceived ease of using the social presence tool was negatively related to students’ success.

Keywords

Personal learning environment, Social interaction, Social presence, Online social presence, Academic success

Introduction

In recent years, the influence of social media has spread into various fields, including education. In particular, researchers have focused on exploring whether social media can provide pedagogical benefits for improving the academic success of students (Junco & Mastrodicasa, 2007; Junco, Heiberger, & Loken, 2010; Junco & Cotten, 2011). Additionally, they have examined the impact of using social media on non-academic skills, such as self-expression, communication and teamwork (Junco, Heiberger, & Loken, 2010), where social interaction plays a crucial role.

Lately, we have also witnessed the spread of personal learning environments (PLEs) (Attwell, 2007) that combine different tools based on social software that supports online learning and provides learners with the opportunity to adapt the learning environment to their learning needs. Chatti et al. (2010a, 2010b) define PLEs as encompassing tacit knowledge nodes, i.e. people, and explicit knowledge nodes, i.e. information.

Although social software used in PLEs contains various solutions, such as software aggregators, it mostly lacks extensibility and portability. As of now, we have not seen the systematic exchange and integration of online presence data from diverse social software tools as a part of modern PLEs (Jovanović, Gašević, & Devedžić, 2009). The main deficiency is in the lack of tight coupling between various tools that influence the online learning process. Consequently, students cannot be fully available and reached by each other whenever they want.

The study described in this paper aimed to meet this deficiency, as we developed a PLE called Online Presence for Learning (OP4L PLE). Within this system, social interaction and social presence were fostered by two communication tools: a social interaction tool and a social presence tool. The main idea was to make students aware of their peers’ online presence regardless of their availability within the PLE. We evaluated a learning strategy based on social interaction and social presence to test whether, and to what extent, learning in the PLE can be improved by utilizing tools. Our main purpose was to draw conclusions that would help us improve the system prototype and develop a successful learning strategy to support social interaction and social presence.
The paper is organized as follows. We start by providing a short background and literature review. Then we describe the systems similar to the one developed in our study and continue by presenting the OP4L PLE. Next, we explain our research questions and present the research methods, procedure and results. The paper ends with related works, conclusions and an outlook for future work.

**Background and literature review**

*Social interaction* refers to a “reciprocal exchange between at least two actors that serves to build relational ties among the actors” (as cited in Walker, 2007, p. 34). Garrison & Anderson (2003) introduced four interactions that occur in educational processes: the interaction between (a) the teacher and learner, (b) the learner and learner, (c) the teacher and content, and (d) the learner and content. Social interaction can also be viewed as a construction of visibility, awareness, and accountability, which are characteristics of so-called *social translucence*. The idea is to “support coherent behavior by making participants and their activities visible to one another” in computer-mediated communications (Erickson & Kellog, 2000, p.59). Similarly, the concept of *social awareness* addresses people’s sense of other people’s social situation and their activities (Gutwin, Greenberg, & Roseman, 1996). In this regard, in computer-mediated communications, and especially within the context of interpersonal social interactions, users can perceive each other more or less “realistically.” This phenomenon is called *social presence*.

Social interaction is closely related to social presence (Tu & McIsaac, 2002). Short, Williams, & Christie (1976) were the first ones to define social presence as the “degree of salience of the other person in the interaction and the consequent salience of the interpersonal relationships” (p. 65). They emphasized the importance of interactions between users within an environment based on its qualities. Later studies (Yen & Tu, 2008) introduced the concept of online social presence as a degree of perception within online communication, as well as feeling within social context, reaction within interactivity and trustworthiness within privacy. Cooke (2007) also added a level of immediacy and intimacy. While immediacy pertains to directness and the intensity of the interaction, intimacy is conveyed through nonverbal communication, such as eye contact, physical proximity and facial expressions.

Both social interaction and social presence have been widely examined within the scope of the question of how they improve online learning. Previous studies (Picciano, 2002; Richardson & Swan, 2003; Swan & Shih, 2005; Russo & Benson, 2005) reported a positive association between perceived social presence and perceived learning, but no connections were found between social presence and grades on the final exam (Picciano, 2002). On the contrary, Liu, Gomez & Yen (2009) reported social presence as “a significant predictor of course retention and final grades in the community college online environment” (p.165). Social presence was also found to be positively related to the students’ perceived satisfaction with web-based learning environments (Richardson & Swan, 2003; Hostetter & Busch, 2006).

Despite the growing interest in using social interaction and social presence in education, no appropriate methodology was found on how they should be effectively adopted into online learning. In existing research, there have been many attempts to address social interaction and social presence in PLEs. Graasp (2013) is a collaborative work platform that supports users in “creating and sharing resources and widgets with other people in the context of space” (Bogdanov, Gillet, & Salzmann, 2013). It enables the aggregation of data and content from different social media and social networking applications. Social interaction and social presence are addressed by allowing users to collaborate and communicate with peers either using the system or external applications for online collaboration or communication. However, Graasp lacks contextualization, which facilitates social presence and is vague about where the system would capture data about users’ learning context and where it would use that data for recommendations.

In contrast, Zheng & Li (2008) introduced a three-dimensional model for facilitating online social presence through the recommendation of peers. It is based on students’ learning context defined as a result of the interaction of three elements: knowledge potential, social proximity and technical access, where account time proximity and technical media are also considered. However, it is disadvantageous not to consider the influence of students’ participation in learning activities based on the decision of whether one is competent enough to be recommended to peers (Jeremić, Milikić, Jovanović, Radulović, & Brković, 2012). In what follows, we present the OP4L PLE developed in our study.
OP4L personal learning environment

The OP4L PLE (2012) is based on both a previously developed learning framework, Design Patterns Teaching Help System (DEPTHS) (Jeremić, Jovanović, & Gašević, 2011) and the Online Presence Ontology Server (OPOS) (OPOS, n.d.). DEPTHS is a PLE that enables collaborative project-based learning and comprises the Moodle Learning Management System (LMS) (Moodle, 2013), as well as several context-aware educational services that provide users with right-in-time learning support adapted to their user profiles and current learning context (Jeremić, Jovanović, & Gašević, 2013). These services use online presence data provided by the OPOS. In addition, OP4L integrates popular social media tools, such as Facebook and Twitter, in order to provide a seamless connection with users who are not momentarily using a learning environment.

Our decision to integrate the above-mentioned tools within the OP4L is based on the popularity of Moodle LMS (integrated within DEPTHS) in educational settings and the popularity of Facebook and Twitter as social media, as well as increased interest among researchers for using social media in educational settings.

In the OP4L PLE, users can learn collaboratively by performing tasks and utilizing various tools in the system. It enables a teacher to design project-based tasks where the users’ learning is supported by social interaction and social presence. While social interaction is technically supported by the Moodle discussion forum, social presence is enabled through the use of the Peers Recommendation Service.

The front end of this service was developed as a Moodle plugin that can be integrated into any learning activity (Jeremić, Milikić, Jovanović, Radulović, & Brković, 2012). Based on the students’ current learning context, it recommends the most relevant peers for communication and resolves their social presence. The selection of peers is performed via a comparison of the learning contexts that other users have been in, as well as by matching their successes and experiences within similar learning contexts (Jeremić, Jovanović, & Gašević, 2013). More precisely, the peers’ relevance is estimated by analyzing three different kinds of knowledge: their estimated knowledge of the topic, their knowledge of similar or related problems and their knowledge of broader topics. Once the service estimates the most relevant peers, it resolves their availability in other social networks and presents this information...
to the end user, so that they become aware of who could be reached by using Moodle or social networks. This module supports the use of the Moodle chat tool and Facebook messaging system. In our study, we use the term social presence tool when referring to utilizing this type of communication tool. The main advantage of integrating social presence tools into the system is providing users with immediate access to their peers, regardless of whether they are using Moodle at the moment or not. It is sufficient for the system if the peers are available on some of the most frequently used social networks. The system alone takes care that they receive the message through the channel connecting Moodle and these social networks.

Figure 1 shows a screenshot of a sample task created by the OP4L module for collaborative work on a common project. A description of the task is given in the main window area, while individual subtasks with additional information are accessible from the table below the description. The student is expected to perform each of these subtasks within the specified time frame. The left-hand side of each window provides access to the recommended peers.

In addition to the Peers Recommendation Service, the environment contains a Semantic Annotation and Indexing Service and a Resource Recommendation Service. These services enhance the learning process through the automatic semantic annotation of internal and external learning sources, such as forum posts, webpages, posted documents and their recommendations within a given learning context (Jeremić, Milikić, Jovanović, Radulović, & Brković, 2012).

The Semantic Annotation and Indexing Service (Jeremić, Jovanović, & Gašević, 2011) is used to index content at public websites and within the system. The module analyzes the text of documents, recognizes specific domain concepts defined in the domain ontology, and finds how relevant it is for a specific domain concept.

The Resource Recommendation Service (Jeremić, Jovanović, & Gašević, 2011) generates a list of recommended online resources or content published within the system, and recommends them to the users based on their current learning context within the PLE. To do this, it crawls and annotates publicly accessible learning resources, computes the relevance of each resource (i.e., Web page) available from these repositories for the student’s current learning context and selects the most relevant pages for the student. The users assess the relevance of the resources and thus improve the overall rating of the resource.

**Research questions**

To outline the benefit of our PLE, in this study we evaluated the users’ academic performance. The main objective was to examine whether and how social interaction and social presence are related to the learning outcome in the PLE, whereas we also considered the students’ motivational orientations and learning strategies.

Figure 2 shows the plan of our investigation with an emphasis on three main elements: social interaction, social presence, and students’ academic success. In particular, we proposed examining the intensity and quality of social interaction, and the intensity and perception of social presence. We focused on the relationships between all three main elements and also suggested the role of motivational orientations and learning strategies in the learning process.

We identified the following research questions to be examined:

**RQ1:** Is there a relationship between the intensity of social interaction and the student’s academic success in the PLE?
We expect to find a positive relationship between these variables, anticipating that the more intensive the students’ social interactions are, the better their success will be.

**RQ2:** Is there a relationship between the quality of social interaction and students’ academic success?
We expect to find a positive relationship between these variables, which reflects that as the quality of social interaction increases, students’ academic success will also improve.

**RQ3:** Is there a relationship between the intensity of social interaction and its quality in a PLE?
A positive association is also expected to be found in answer to the third research question. We assume that as the intensity of social interaction increases, its quality will also increase.
**RQ4:** Is there a relationship between the intensity of social presence and students’ academic success? We anticipate a positive relationship between the intensity of social presence and students’ academic success. The intensity of social presence reflects the intensity of using the social presence tool. Thus, we anticipate that as the use of the social presence tool increases, academic success will improve.

**RQ5:** Is there a relationship between perceived social presence and students’ academic success? We expect to find a positive relationship between these variables. The perceived social presence is determined by the perceived ease of use of the social presence tool. We predict that with the improvement of the perceived ease of use of the social presence tool, that academic success will improve as well.

**RQ6:** Are there statistically significant differences between student groups with a higher intensity of social interaction and student groups with a lower intensity of social interaction in motivational orientations? It is expected that motivational orientations will differ between student groups according to the intensity of their social interactions. We assume that the members of a student group with a higher intensity of social interaction will report higher mean scores in motivational orientation than students with a lower intensity of social interaction.

**RQ7:** Are there statistically significant differences in learning strategies between student groups with a higher intensity of social interaction and student groups with a lower intensity of social interaction? In learning strategies we expect to find statistically significant differences, where student groups with a higher intensity of social interaction will report higher mean scores than groups with a lower intensity.

---

**Figure 2.** Social interaction, social presence and academic success in interrelation with motivational orientations and learning strategies

---

**Methods**

**Participants**

The participants in the experiment were 62 students of electrical engineering from the University of Maribor in Slovenia. All students were male, with an average age of 20 years (age range: 19–25, SD = 1.41). Student volunteers were selected based on their previous experience with the programming language C. Before the experiment, they attended a few-hour training course where the main features of the system were presented. We randomly classified
them into three groups with 8 students, two groups with 9 students and two groups with 10 students. The sizes of the groups were based on our teaching experience to assure an effective educational process.

Measures

The measuring instruments were classified into two groups: (1) questionnaires and (2) server log files and academic success. We used questionnaires to collect responses about social presence, motivational orientation, learning strategies and participants’ profiles. We also analyzed a dataset of the server log files and academic success to define the intensity and quality of social interaction. In this way the results from the questionnaires were elaborated with research findings from the server log files analysis.

Questionnaires

To obtain data about the participants’ perception of their experience with the OP4L PLE, we applied a quantitative research method (a questionnaire survey). The following questionnaires were used: (a) Social Presence Questionnaire, (b) Motivated Strategies for Learning Questionnaire, and (c) Profile Questionnaire. A questionnaire survey is available on the following address: http://medijske.um.si/doc/OP4L_questionnaire_ENG.pdf.

- Social Presence Questionnaire. The purpose of this questionnaire was to measure students’ performance when using the social presence tool and consequently to examine subjective satisfaction of the users as one of the aspects of usability (Holzinger, 2005). The questionnaire encompassed eight questions classified into two sections: (a) the intensity of using the social presence tool and (b) the students’ perceived ease of using the tool. The first section comprised two multiple-choice questions and the second section comprised six close-ended questions. The questions in the first section recorded the frequency of using the social presence interaction tool, regarding two communication types: student-to-student and student-to-teacher communication. The participants responded with answers ranging from 1 (never) to 5 (very regularly – I communicated every time I encountered a problem when completing a task in the OP4L PLE). The second section measured the students’ perceived ease of using the social presence tool. The participants responded with answers ranging from 1 (strongly disagree) to 5 (strongly agree). The alpha reliability coefficient (Cronbach, 1951) of the questions was .81.

- Motivated Strategies for Learning Questionnaire (MSLQ). The MSLQ questionnaire was applied to measure students’ motivational orientations and their use of learning strategies (Pintrich, Smith, Garcia, & McKeachie, 1991). We used 81 statements administered with 7-point Likert-type response categories ranging from 1 (strongly disagree) to 7 (strongly agree). Measuring motivational orientations comprised sets of questions within the following components: intrinsic and extrinsic goal orientation, task value, control of learning beliefs, self-efficacy for learning and performance, and test anxiety. Measuring learning strategies contained the following components: rehearsal, elaboration, organization, critical thinking, metacognitive self-regulation, time/study environmental management, effort regulation, peer learning and help seeking (Pintrich, Smith, Garcia, & McKeachie, 1991).

We used the MSLQ questionnaire to identify the students’ sources of motivation and learning strategies used during the learning process, as well as to find relations with the students’ learning activities in the PLE and the students’ academic success.

- Profile Questionnaire. The Profile Questionnaire was applied to gather data about the participants’ gender, age and programming knowledge. The level of knowledge of the programming language C was measured with one 5-point Likert-type item with response categories ranging from 1 to 5. The students identified their knowledge on a scale ranging from beginners (response category 1), who had no experience in programming, to experts (response category 5), who had a solid knowledge of the programming principles in C. In addition, we asked students one open-ended question regarding their personal opinions about collaborative learning in the OP4L PLE.
Server log files and academic success

We obtained data about the students’ social interactions and academic success. The OP4L PLE has a tracking mechanism that collects and stores the users’ traces in a database from which all the necessary data with respect to the aim of this study could be extracted.

- Dataset of Social Interaction. We measured two separate elements: the intensity and the quality of the social interaction. The intensity was measured as an average number of discussion forum posts per student in each group.

  The quality of the social interaction was defined by the teacher’s assessment of students’ discussion forum posts in the OP4L PLE. The teacher assigned a mark on a scale of 0 to 5 for each post. The mark 0 was assigned if no post was published by the student and the mark 5 was assigned to an excellent post. During the assessment, two different categories of forum posts were considered. The first category included the discussion forum posts, where the students triggered a discussion via a request for assistance or additional explanation, or a question raising a dilemma, proposal or suspicion. Each post was assessed by the teacher by considering the significance of the question raised by the student, according to the difficulty of the given task and providing another possibility of the question being answered by other students and by the teacher.

  The second category comprised discussion forum posts where students responded to questions raised in posts defined within the first category. The assessment of these posts was also based on the significance and technical correctness of the posts from the viewpoint of the task content; the factor of predicting how relevant the post was for its ability to guide the user to a proper solution was considered as well.

  The teacher’s marks for the discussion forum posts were summed up for each group of students. The results were divided by the number of posts in the group. Then the average mark of posts per group was computed. The final results varied on a scale of 0 to 5.

- Dataset of Students’ Academic Success. The students’ academic success was measured in the teacher’s final grade, assessing the quality of the submitted projects. The grades were assigned on a scale of 0 to 5. The grade 0 was assigned when the project was not submitted or the quality of the submitted project was assessed as being below 50%. The grades increased proportionally with percentages. The grade 5 was assigned to submitted projects that were assessed at 91% or more.

Procedure

The experiment was conducted in May 2012 at the University of Maribor in Slovenia and was performed in three steps: (1) demonstration of using the system along with a training session, (2) working with the OP4L PLE, and (3) assessment of the OP4L PLE.

In the first step, we demonstrated the system’s functionalities during a training session where an example of the assignment was introduced. The participants were informed about the evaluation criteria to become aware of the teacher’s expectations regarding their activity in the OP4L PLE.

The second step included the main experiment. The teacher gave each group its own assignment. Although all members within each group received the same assignment, they were asked to submit their own solutions. The assigned task was complex and challenging, so that it could elicit a constructive learning process in students (Van Merrienboer & Paas, 2003). Learning process design in PLEs has its roots in various theories. The proposed opportunities for communication are related to the Social Learning Theory (Bandura, 1977) and collaboration (Dillenbourg, Baker, Blaye, & O’Malley, 1994), referenced together as the Social Interaction Learning Theory, where social interaction is a crucial element in learning.

The learning process encompassed three predefined sets of tasks distributed evenly over a three-week period:
1. brainstorming and submitting an idea of the solution, where the participants shared their ideas about a solution with other participants in the discussion forum, used recommended resources and communicated via the social presence tool,
2. submission of the solution: programming and submitting the program, where the participants wrote a computer program in C, discussed it on the forum, used recommended resources and finally submitted the project.
3. evaluating each other’s submitted solution, where the participants evaluated their own and each other’s submitted project.

In the third step, the participants filled in the questionnaires to assess the OP4L PLE regarding their learning experiences within this learning environment. Throughout the course, the teacher was fully engaged in the students’ learning process and available 24 hours a day.

Statistical analyses

The internal consistency reliability of a set of items for one variable was checked with the Cronbach's Alpha coefficient (Cronbach, 1951). Statistical associations between variables were inspected with parametric correlation (Pearson's correlation coefficient). Statistically significant differences between independent samples were examined with a One-way Analysis of Variance (One-way ANOVA) statistical model (Howell, 2002). All analyses were performed using SPSS version 20.0 software.

Results

Detailed analysis

The first three research questions address the relationships between (a) the intensity of a social interaction, (b) the quality of a social interaction, and (c) the students' academic success. The intensity was measured by the numbers of students' discussion forum posts in the OP4L PLE within the group; the quality was measured by the teacher's assessment of the students' forum posts and academic success was measured by the teacher’s assessment of the students' submitted projects.

Table 1 shows the mean scores of these variables for student groups, using descriptive statistics. The students in group 6 reported the most intensive social interaction (mean (M) = 3.5) and also reached the highest quality level of social interaction (M = 3.86, standard deviation (SD) = .73). Comparing the scores of academic success, student group 2 reported better scores (M = 4.50, SD = .76) than group 6 (M = 4.20, SD = .63).

A bivariate analysis was conducted to further investigate the first three research questions. A statistically significant positive relationship was found between the intensity and the quality of the social interaction, \( r = .37, \ p < .05 \). It indicates that as students were more active in discussion forum postings in the OP4L PLE, the quality of their posts improved. Likewise, a statistically significant relationship was demonstrated between the quality of social interaction and academic success, \( r = .48, \ p < .05 \), meaning that as the quality improved, their academic success also improved.
In addition, the intensity of the social interaction was significantly correlated with academic success, \( r = .44, p < .05 \). The result indicates that as the number of forum posts increased, the level of success also improved. Figure 3 shows the mean scores of the three variables in student groups.

![Graph of mean scores for intensity of social interaction, quality of social interaction, and academic success in student groups](image)

**Figure 3.** Graph of mean scores for (a) intensity of social interaction, (b) quality of social interaction and (c) academic success in student groups

Considering student groups with different levels of intensity with regard to using the social interaction tool, we conducted an analysis with one-way ANOVA to detect statistically significant differences between student groups in all three variables. The results demonstrated statistically significant differences between the groups in the variable “intensity of social interaction,” \( F(6, 156) = 9.98, p < .01 \). Likewise, statistically significant differences were found in the variable “quality of social interaction,” \( F(6, 156) = 11.86, p < .01 \), and “students’ academic success,” \( F(6, 55) = 4.46, p < .01 \). However, the result did not indicate whether the effect was reported within or between the groups. Thus, post hoc comparisons using the Tukey HSD test (Howell, 2002) were conducted. Statistically significant differences were found between the student groups 5 and 6 in the variables intensity, quality of social interaction and academic success.

The fourth research question asked if there was a relationship between the intensity of social presence and academic success. The results of the analysis showed no statistically significant associations, \( r = – .15, p > .05 \). The fifth research question pertained to the association between perceived social presence and success. The results revealed a negative relationship between these variables, \( r = – .26, p < .05 \), meaning that as the students’ satisfaction with utilizing a tool increased, academic success decreased.

The sixth research question asked whether there were statistically significant differences between student groups with a higher intensity of social interaction and student groups with a lower intensity of social interaction in motivational orientations. The same differences were addressed in the seventh research question for learning strategies. To find answers to both research questions, we examined the students’ motivational orientations and learning strategies. Based on a set of questions in the MSLQ questionnaire (Pintrich, Smith, Garcia, & McKeachie, 1991), we identified the motivational orientations and learning strategies that were used by the students. Figure 4 shows to what extent the motivational orientations were applied by students.

As one can see from Figure 4, among motivational orientations, the highest mean score was detected in the variable “control of learning beliefs” (\( M = 5.75, SD = 0.93 \)). This variable illustrates the students’ beliefs that they will gain positive benefits from learning through performing a task (Pintrich, Smith, Garcia, & McKeachie, 1991, p.12). The variable “test anxiety,” defined as an unpleasant feeling or emotional state toward tests, was assessed with the lowest mean score among motivational orientations (\( M = 4.05, SD = 1.22 \)).
Among motivational orientations, statistically significant differences were only found between student groups 5 and 6. Two types of motivational orientations reported significant results: intrinsic goal orientation and self-efficacy for learning and performance. Intrinsic goal orientation reflects the fact that students are curious about the assigned task and feel it to be a challenge. Self-efficacy illustrates how students judge and assess themselves and their abilities at performing a task (Pintrich, Smith, Garcia, & McKeachie, 1991).

Statistical differences were checked by conducting a one-way ANOVA. In intrinsic goal orientation, statistically significant differences were detected between the groups 5 and 6, $F(6, 55) = 4.13, p < .01$. The mean score for the variable “intrinsic goal orientation” was significantly lower in group 5 ($M = 5.06, SD = .75$) than in group 6 ($M = 6.35, SD = .47$). In self-efficacy, statistically significant differences were also detected between the groups 5 and 6, $F(6, 55) = 2.43, p < .05$. Likewise, the mean score for the variable “self-efficacy for learning and performance” was lower in group 5 ($M = 4.94, SD = .47$) than in group 6 ($M = 5.91, SD = .64$).

Figure 5 shows that among learning strategies, the highest mean score was found in the learning strategy “organization” ($M = 5.27, SD = .93$), which pertains to students’ selection of pieces of information and connecting them into the whole to be learned (Pintrich, Smith, Garcia, & McKeachie, 1991, p. 21). The lowest mean score was
detected in the learning strategy “critical thinking” \((M = 4.31, SD = .97)\). It indicates the students’ application of previously gained knowledge to the new context in order to be able to find solutions to new problems (Pintrich, Smith, Garcia, & McKeachie, 1991, p. 22). In learning strategies, no statistically significant differences between student groups were detected.

In the profile questionnaire, the students expressed their opinions about collaborative learning. The answers showed that the majority of them were satisfied with the method and organization of learning. As the main advantage that was indicated was access to different opinions in one PLE. Participants also noted their criticisms, which were primarily directed at technical issues. For instance, some participants would have preferred improvements in the user interface, a simplified communication process by using the social presence tool and improvements in overall system responsiveness.

**Summary**

Table 2 summarizes the answers to our seven research questions. The study was aimed at contributing to the field of students’ learning and communication experiences in PLEs. The purpose was to investigate how social interaction and social presence are related to students’ academic success and how motivational orientations and learning strategies interplay with them. To meet this aim, we integrated the social interaction and the social presence tool in a PLE, and conducted an experiment.

<table>
<thead>
<tr>
<th>RQ</th>
<th>Variable 1</th>
<th>Variable 2</th>
<th>Observed relationship</th>
<th>Observed difference</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQ1</td>
<td>intensity of social interaction</td>
<td>academic success</td>
<td>x</td>
<td>(r = .44, p &lt; .05)</td>
<td></td>
</tr>
<tr>
<td>RQ2</td>
<td>quality of social interaction</td>
<td>academic success</td>
<td>x</td>
<td>(r = .48, p &lt; .05)</td>
<td></td>
</tr>
<tr>
<td>RQ3</td>
<td>intensity of social interaction</td>
<td>quality of social interaction</td>
<td>x</td>
<td>(r = .37, p &lt; .05)</td>
<td></td>
</tr>
<tr>
<td>RQ4</td>
<td>intensity of social presence</td>
<td>academic success</td>
<td>x</td>
<td>(r = -.15, p &gt; .05)</td>
<td></td>
</tr>
<tr>
<td>RQ5</td>
<td>perceived social presence</td>
<td>academic success</td>
<td>x</td>
<td>(r = -.26, p &lt; .05)</td>
<td></td>
</tr>
<tr>
<td>RQ6</td>
<td>higher intensity of social interaction in</td>
<td>lower intensity of social interaction in</td>
<td>x</td>
<td></td>
<td>intrinsic goal orientation:</td>
</tr>
<tr>
<td></td>
<td>motivational orientations</td>
<td>motivational orientations</td>
<td></td>
<td></td>
<td>between the group 5 and 6, (F(6, 55) = 4.13,)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(p &lt; .01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>self-efficacy:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>between the groups 5 and 6, (F(6, 55) = 2.43,)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(p &lt; .05)</td>
</tr>
<tr>
<td>RQ7</td>
<td>higher intensity of social interaction in</td>
<td>lower intensity of social interaction in</td>
<td>x</td>
<td></td>
<td>no statistically significant differences</td>
</tr>
<tr>
<td></td>
<td>learning strategies</td>
<td>learning strategies</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Conclusions and future work**

In this study, we found that the intensity and quality of a social interaction are connected to students’ academic success. However, no connections were found between social presence and students’ academic success. These findings are partly in alignment with the findings of Picciano (2002) who found no connection between perceived social presence and final exam grades; conversely, he found a strong relationship between social presence and students’ performance on written assignments. The reason can be in supporting social presence on a different level. In contrast to our study, Picciano (2002) supported social presence in the learning environment on a communication level, such as using names, etc., but not with a social presence communication tool, as we did. Additionally, our
findings substantiated that it is not the intensity of social presence, but rather the intensity of social interaction that is connected to students’ academic success.

Our findings provide clues as to how to carefully design and develop communication tools to ensure a positive impact on academic success, such as providing various communication tools that actively involve social media throughout the learning process. We also suggested a model to technically implement them into the learning process to support pedagogical activities. For communication tools integrated into PLEs, it is important that the designers are convinced that the users perceive the tool as easy to use, which can lead to more intense use and ultimately induce a positive relationship with the learners’ academic success. We suppose that our users experienced a lack of perceived ease of use of the social presence tool, which may have influenced the user experience.

The limitation of our study stems, first, from the particularity of the sample, including the fact that it exclusively consisted of male electrical engineering students. The results would probably differ if females and/or social science students were included in the study, as the nature of the task would be different. Second, our study was only limited to the Moodle LMS, and not Blackboard (2013), OLAT (2013), Edmodo (2013) or other e-learning platforms that might impact users’ performance when using services.

Regarding future work, there are several directions to be investigated. Firstly, the findings of this study will help us improve the system prototype and integrate an improved social interaction and social presence tool more efficiently into the PLE to ascertain the use of both tools in the field. Our prototype uses a set of tools that we believe are often used in an academic environment. However, all the tools are developed in such a way that they could be easily integrated into other e-learning systems. Our intention is to investigate and experiment with different e-learning systems, settings and social groups in order to verify the success of this approach.

Secondly, in that regard, investigating the integration of Twitter in the social presence tool might be relevant due to the distinction between Facebook as a social networking site and Twitter as a micro blog. While Facebook is more directed at social connections, Twitter is more about staying informed. Integrating both tools may thus enable researchers to compare the differences in users’ perception of using both tools in the PLE.

Acknowledgements

This study was part of the project OP4L (SEE-ERA.NET PLUS no. 115) and was supported by SEE-ERA.NET PLUS Coordination and Support Action of the European Community. The study was also supported by the Slovenian Research Agency [1000-11-310140].

References


Exploring the Relationship between Sanctioned and Unsanctioned Laptop Use in a 1:1 Classroom

Martin Tallvid¹*, Johan Lundin¹, Lars Svensson² and Berner Lindström¹

¹University of Gothenburg, Sweden // ²University West, Sweden // martin.tallvid@ait.gu.se // johan.lundin@gu.se // lars.svensson@hv.se // berner.lindstrom@gu.se

*Corresponding author

(Submitted September 29, 2013; Revised February 5, 2014; Accepted February 18, 2014)

ABSTRACT

The research reported in this article explores and discusses students’ use of laptops in a 1:1 setting. A problem experienced by teachers is that the laptops are possible distractors and tempt students to engage in use that is not in line with the teacher’s idea of what would be suitable in relation to the current assignment. Three annual surveys in combination with interviews and classroom observations were carried out in two Swedish secondary schools during a phase of the implementation of 1:1-laptops. The results show that there is not a reciprocal correlation between sanctioned laptop use and unsanctioned laptop use. The findings also show that the students’ unsanctioned use of laptops was relatively high, but stable throughout the duration of the three years. Furthermore, results show that the number of students who do not game or chat at all has increased every year. The findings have implications for the discussions concerning the use of personal laptops in secondary schools.

Keywords

Laptop, 1:1, Education, Computer, K-12

Introduction

“Surveillance software? No way! We’ll take the discussion – if it is needed.”
“Problems with unsanctioned use? Maybe they occasionally do things I don’t want them to do. But, on the other hand, I think they did that prior the laptop as well.”

(Excerpts from an interview with a secondary school teacher)

In a 1:1-classroom (i.e., the students are each provided with a laptop), teachers are confronted with constant tension between two parallel agendas: either promoting laptop use that is desired and sanctioned, or preventing laptop use that is not welcome nor sanctioned (Fried, 2008; Kay & Lauricella, 2011; Trimmel & Bachmann, 2004). Results from earlier research revealed a preconceived belief among teachers that there is a reciprocal relationship between the two agendas (Tallvid, 2010). The idea originates from the presumption that the more students engage in activities they are not supposed to, the less they are likely to harvest the potential benefits from using the laptop the way they should and vice versa. However, there is no earlier research that can confirm or confute this relationship.

We will report on a study of how students in two Swedish secondary schools used their laptops in their everyday educational activities. The use of laptops in these schools is regulated by discussing norms for use, rather than by introducing formal rules or technical restrictions. The data does not focus on laptop use as being part of specifically designed educational tasks, but rather on the everyday use of the laptops. This includes taking notes, or listening to music during class, as well as use that is not sanctioned by the teachers, such as playing games or chatting.

The purpose of the study was to examine how a 1:1 laptop initiative in two schools affected student use of laptops. Our intention was not to evaluate a particular technology or method for using ICT, nor to investigate if students learn more or less with their mobile devices. In our perspective, the laptop and other tools in a classroom are culturally and historically situated, and thus we need to understand and study this context to appreciate the circumstances surrounding the use. In such a perspective, the laptops are tools and are part of what constitutes a social practice. Therefore, questions about the character of the tools, how they are employed in practice, what knowledge is needed to be able to use them, how such knowledge is learned, and how they need to be negotiated into the day-to-day practice are of interest. Hence, the methods to study the implementation of ICT in schools have to be adapted. We consider it essential to conduct studies of day-to-day use of ICT, in order to understand the consequences of the use (Selwyn, 2011). Experimental studies focusing on the introduction of new technologies certainly have a place, but we also need to look more into the non-experimental, routine, and present use of technology by students. Such an
Two research questions were investigated.
• What is the relationship between sanctioned and unsanctioned laptop use in a 1:1 classroom, where use is regulated by discussing norms for use, rather than by introducing formal rules or technical restrictions?
• How does the students’ use of laptops in a 1:1 classroom change over time?

Laptops in the classroom - relevant research

Laptops connected to the Internet offer several opportunities for use in the classroom. Potential benefits and pitfalls have been studied since laptops were introduced in educational settings, and educational reformers have presented wireless laptops as the next great educational innovation (e.g., Brown & Petitto, 2003). However, laptops have also been described as a possible reason for a decreased academic performance (Hembrooke, 2001). Apart from how the use of laptops relates to learning a particular content, it has also been suggested that laptops promote collaborative working-methods and project-based instruction, and also develops the students’ skill in the use of technology (Smart, Kumar, & Kumar, 2004). It has also been described how an implementation of laptops were affected by culture and gender (Saunders & Quirke, 2002). While many of these earlier efforts focused on higher education, we now see how laptops are becoming increasingly important in K-12 settings, in particular in so-called 1:1 programmes.

1:1 programmes are mainly motivated by the possibilities to improve student learning in general, and to prepare them for participation in the knowledge society; sometimes articulated as 21st century skills (Cogan-Drew, 2010; Rosefsky & Opfer, 2012). Research shows how laptop programmes have positive effects on student achievements in general, and how laptop use enhances learning and promotes interactions between students (e.g., Barak, Lipson, & Lerman, 2006; Bebell & Kay, 2010; Warschauer, 2006). The use of laptops in the classroom can increase students’ motivation, and their ability to gain understanding, and can also increase their overall educational achievements (Samson, 2010; Wurst, Smarkola, & Gaffney, 2008). In comparative studies, students involved in 1:1 programmes improved their educational outcomes, compared to students without personal laptops (e.g., Bebell & Kay, 2010; Bebell & O’Dwyer, 2010; Zucker & McGhee, 2005).

In contrast to the research that emphasizes the benefits of 1:1, some research has also reported on the setbacks. Rather than questioning the potential benefits of using the laptop for educational activities, these studies focus on the unwanted activities. The main problem experienced by teachers is that the laptops offer distractions and tempt students to engage in use that is not in line with the teacher’s idea of what would be suitable in relation to the current assignment (e.g., Aguilar-Roca, Williams, & O’Dowd, 2012; Fried, 2008; Gaudreau, Miranda, & Gareau, 2014; Kay, 2012; Kraemer, Dedrick, & Sharma, 2009; Lauricella & Kay, 2010; MacKinnon & Williams, 2006; Reynol, 2012; Weston & Bain, 2010; Wurst et al., 2008; Yamamoto, 2007). Hu (2007) reported on schools abandoning 1:1 initiatives due to misuse of the laptops by the students: cheating on tests and playing on-line games during class. Some would call this “off task-use” (Gulz, Silvervarg, & Sjödén, 2010; Hofer, 2007; Wood, Zivcakova, Gentile, Archer, & De Pasquale, 2012), that is, use that is considered to break classroom norms and to have negative consequences on students’ learning (Guribye, 2005). Conflict concerning the use of laptops in classrooms is not surprising. When introducing new technologies into educational settings, or for that sake into any practice, there will be conflicting views on the benefits of the new tools, and a need for negotiation of how these tools are to be used (Lundin, 2005).

However, as proposed by Mifsud and Mörch (2010), a dichotomous view on suitable use of the laptop is problematic. Apart from clearly abusive use, there are a number of less easily defined uses and they suggest that the term “off task” should be reconsidered. A range of research shows that learning activities often are multifaceted and that it is complicated to pinpoint what can be considered to be an “on task” or “off task” activity (Björkval & Enblom, 2010; Cismaru, 2011; Fried, 2008; Maybin, 2007). Tallvid, Lundin and Lindstrom (2012) exemplify how the students sometimes elaborate tasks given by the teacher. The elaboration is not always within the teacher-defined task design, but it can be inside the curriculum and learning goals of the student.
This has led to discussions regarding if and how students’ use of laptops can be monitored and regulated by rules and/or software filtering (Owen, Farsaii, Knezek, & Christensen, 2006; Warschauer & Grimes, 2005). Due to the temptations offered by the wireless, Internet-connected laptop, and, in particular, how these can distract the individual student or disturb the collective use, some teachers tend to restrict or ban the use of laptops in the classroom (Young, 2006). Activities that are understood as problematic and disturbing to teachers are by no means unfamiliar to scholars or teachers. Classroom discipline has always been a problem, and research shows that teachers repeatedly report that teaching is disturbed by students’ disruptive behaviour in general (Akiba, LeTendre, Baker, & Goesling, 2002; Veenman, 1984).

The case setting

The data in this article is part of a larger data set collected over a three-year period during a 1:1 project in two schools. The laptop programme was introduced in 2007 and was the first, large-scale initiative with 1:1 in municipal schools in Sweden. The implementation phases have been described in evaluation reports (Hallerström & Tallvid, 2008; Tallvid, 2010; Tallvid & Hallerström, 2009). The initiative for the 1:1 project was taken by the two headmasters in the participating schools, together with two representatives from the municipality. Together they formed a steering group and set up goals for the project. This article focuses on laptop used by students that have unrestricted access to Internet in the classrooms. The steering group made a recommendation concerning their use that was clear and concise: the students were responsible for their own use. No software filters should be installed and there should be no restrictions, with the exception of illegal activities, to the use of the Internet. The overall attitude and rule was articulated as: “the filter should be in your mind – not inside the laptop” (Tallvid, 2010, p. 14). Both schools had wireless networks and unrestricted access to the Internet. The students had access to their laptops twenty-four hours a day, even during weekends and holidays. They were not allowed to leave the laptops at school and they had to take responsibility for keeping the computers updated and for making their own back-ups.

During the initial phase of the 1:1 project, some teachers expressed doubts concerning the ways the students were using the laptops. Two of the teacher-teams (one at each school) were not content with the “absence of rules postulated by the steering-group” (quote from interview with teacher) and tried to introduce their own, more restrictive, rules and guidelines. They wanted to have the possibility to confiscate the laptop if the student used it for playing games during class, for illegal file sharing or for downloading copyright-protected material. The steering group reacted strongly on this attempt to establish restrictions, and instead of allowing the teachers to ban the laptops or to introduce stricter rules; they urged the teachers to discuss the ethical questions more intensely in the class. The steering group emphasized that the laptops should be considered as tools for learning and consequently, should not be removed from the students. Apart from this occasion in the initial phase of the 1:1 implementation process, the teachers had only minor concerns about rules for laptop use.

The students used the laptops on a daily basis. As in most Swedish secondary school classrooms, the teaching methods were a mixture between teacher-centered and student-centered approaches (Håkansson & Sundberg, 2012).

Method

Data from 500 students and 60 teachers was collected through an annual web-based questionnaire (2007 – 2010), supplemented by interviews with teachers, students and representatives of the steering group, and by regular observations in the classrooms (60 hours). To make it possible to distinguish changes over time, the selected data used in this article originate from the same group of students over the three years, starting in 7th grade (year one) and ending in 9th grade (year three).

The interviews with the students were semi-structured and performed in groups of eight. The interviews were digitally recorded and notes were taken. The field notes were written shortly after the interviews. Classroom observations were conducted weekly during the three years of the 1:1 project and four observations were video recorded. The first survey was conducted after three months of laptop use, when the respondents were in 7th grade. The two following surveys were conducted during spring-term when the respondents were in 8th grade and 9th grade.
The questionnaire investigated a set of demographic variables (e.g., gender, age, and school), respondents’ attitudes to using laptops, and the nature and frequency of laptop use. The categories were derived from analysis of the initial classroom observations.

An annual survey was conducted to measure the different types of laptop use. The questionnaires to teachers and students had an identical core, with a few minor exceptions and additions.

<table>
<thead>
<tr>
<th>Year 3 (2010)</th>
<th>Sampled n = 137</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response rate = 76%</td>
<td></td>
</tr>
<tr>
<td>Year 2 (2009)</td>
<td>Sampled n = 142</td>
</tr>
<tr>
<td>Response rate = 80%</td>
<td></td>
</tr>
<tr>
<td>Year 1 (2007)</td>
<td>Sampled n = 114</td>
</tr>
<tr>
<td>Response rate = 67%</td>
<td></td>
</tr>
</tbody>
</table>

7th grade 8th grade 9th grade

The survey investigated the frequency of use (Many times every day, Occasionally every day, Occasionally every week, Occasionally every term, and Never/almost never) of the laptops for the following activities: information search, word processing, downloading films or music, chatting (without permission from teacher), playing games (without permission from teacher), surfing on Internet (without permission from teacher), preparing or performing presentations, sound recording, listening to music, and using the laptop’s camera.

**Results**

Based on interviews with teachers and classroom observations it was possible to determine two main categories of laptop use: “sanctioned use” and “unsanctioned use”. The first use-category contains activities that are allowed by the teacher and were usually initiated by the teacher, for example, activities in which students were asked to use the Internet to find facts and prepare a digital presentation or in which they were asked to write an essay and send it to the teacher. This category also included use that was not initiated by the teacher. Occasionally, students used the laptop for listening to music without an explicit suggestion from the teacher. This was widely accepted during individual work, such as reading and doing mathematics. These kinds of uses are defined as “sanctioned use”.

In contrast, the second use-category was clearly neither accepted nor tolerated by the teachers because the activities were either noisy or in other ways challenging the prevalent norms of how students should behave in a classroom. For example, it was clearly against the norms to engage in playing games or browse the web without a relevant educational purpose. These kinds of use are defined as “unsanctioned use”. Table 2 provides a list of indicators of sanctioned and unsanctioned laptop use.

Some of the use-categories demanded a certain level of activity and engagement, whereas other categories required less engagement from the students. Furthermore, the observations in the classrooms made it possible to study the teachers’ different reactions concerning the students’ laptop use.

<table>
<thead>
<tr>
<th>Use category</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sanctioned use</td>
<td>Use of word processor, Information searching, Preparing presentation*, Digital recording*, Music listening (using earphones).</td>
</tr>
<tr>
<td>Unsanctioned use</td>
<td>Chatting, Playing games, Off-task surfing*, Downloading.</td>
</tr>
</tbody>
</table>

Only measured in 8th and 9th grades

Many of the students' laptop activities in a 1:1 classroom occur either in convergence with the common teacher-formulated activity or in divergence from these activities, thus distracting attention from them (Brown, McGregor, & Laurier, 2013). Furthermore, many of the activities occur simultaneously and consequently, it is difficult for the teachers to monitor them, even though teachers might have different levels of acceptance. Observations and interviews revealed a shared understanding among teachers at both schools concerning what types of laptop use teachers consider as counter productive to learning.
Sanctioned use

The sanctioned use of the laptops increased during the three years. At the introduction of the 1:1 project, less than half of the students (43%) used the laptop on a daily basis to search for information. By the end of the project, the daily use had increased to 90%. The use of a word processor to take notes and work on assignments had a very similar development, from low use to daily use in three years.

Not surprisingly, 90% of the students use the laptops for information search and word processing every day. These results correlate well with earlier research (OECD, 2009; Pelgrum & Law, 2003; Penuel, 2006), and are activities used in tasks initiated and sanctioned by the teacher.

In Figure 3, we present the sanctioned laptop use as reported in the final survey (9th grade). After three years, information searching and use of word processors are the dominant fields of laptop use, whereas most students did digital recording of sound and took photos/film occasionally. The use of the laptop as a tool for assignments to make presentations or demonstrations seems to be very common, as almost 60% of the students prepared presentations on daily/weekly basis.

The data from the final survey have also been examined for correlational relationships between different user patterns (see tables 4 and 7).
Unsanctioned use

Unsanctioned use includes web surfing as well as downloading music or movies, playing games and online chatting, all being potentially disturbing activities. Playing games and chatting was surveyed over the three years, web surfing only over the two last years.

Students’ game playing was an often-discussed topic among the teachers. The interviews revealed different opinions, including whether it should be allowed at all. If game playing were to be allowed, the arguments concerned what different types of games that should be sanctioned at school. Discussions were common concerning whether it was feasible for teachers to teach about behaving in an ethically correct manner, and subsequently to allow students to play violent war games.

![Figure 4. How playing games changed from 7th to 9th grade](image)

Figure 4 shows that the proportion of the students who never played games during class increased over the years (from 23 % in 7th grade to 45 % in 9th grade), while a stable proportion (approx. 25 %) of the students continued to play games daily over the years. The increasing percentage of students that never played games comes mainly from the group of students that played occasionally. This group decreased from 49 % in 7th grade to 29 % in 9th grade. It is worth noticing that it was asked for how many times a day the students engaged in playing games, not the amount of time playing games took.

![Figure 5. Frequency of playing games during class (9th grade), divided by gender](image)

Figure 5 shows how, in the ninth grade only, just over 13 % of the students played games several times during the school day. It also becomes apparent that most of the playing games is done by boys. In fact, out of the boys almost half (47 %) played computer games daily during class. According to recently published statistics from Sweden (Findahl, 2012), digital playing games differs substantially between boys and girls when it comes to Internet-use at home. However, 37 % of boys between the ages of 12-25 play games on a daily basis, compared to 11 % of the girls.
In our study, a similar pattern seemed to be maintained at school, but the difference is even more significant. Only 5% of the girls played games every day during class, but almost 30% of the boys did. Noticeable is the increase in the number of students that did not play games at all. In the first year, 23% of the students in 7th grade reported that they never played games during class, while in the 8th grade, this rose to 36%. In the last year, 44% of the same group reported that they never played games during class.

![Chat in class](image)

Figure 6. How chat changed from 7th to 9th grade

Figure 6 shows the results for chatting during class. Chatting shows a trend similar to playing games. Around half of the students chat during class (which is a higher proportion than those who play games during class). At the same time, the percentage of students that never chat during class increases over the three years (from 11% in 7th grade to 26% in 9th grade). The increasing percentage of students that never chats comes mainly from the group of students that chats occasionally. There were no noteworthy differences between boys and girls concerning chatting during class.

**Passive use**

Two of the use-categories mentioned in Table 2 stand out from the others: “Music listening” and “Down loading.” Firstly, listening to music, which was not initiated nor always officially encouraged by the teachers, was still allowed as long as the students used earphones and did not disturb their peers. Listening to music was a very common activity. Just above 80% of the students were engaged in this on a daily basis. As soon as there were opportunities for students to work on their own, for example, when studying mathematics, they used the laptop as a music player. Secondly, they used the laptops as a tool for downloading files, such as music or movies, an activity that goes on mainly in the background and does not require any attention from the student. Even though this use was not disturbing, nor did it interfere with other activities, it was still not allowed by the teachers. These kinds of activities are performed while doing other things and can be either sanctioned or unsanctioned. We define them as "passive use." As described below, there are small, but significant, correlations, mainly between the unsanctioned activities.

**Table 3. Correlation between music listening/downloading and use-categories**

<table>
<thead>
<tr>
<th>PEARSON CORRELATION</th>
<th>Music listening</th>
<th>Down loading</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Information search</td>
<td>.233**</td>
<td>-.082</td>
</tr>
<tr>
<td>2. Word processing</td>
<td>.162</td>
<td>-.140</td>
</tr>
<tr>
<td>3. Presentations</td>
<td>.148</td>
<td>.163</td>
</tr>
<tr>
<td>4. Digital recording</td>
<td>.126</td>
<td>.248**</td>
</tr>
<tr>
<td>5. Playing games</td>
<td>.179*</td>
<td>.265**</td>
</tr>
<tr>
<td>6. Web surfing</td>
<td>.387**</td>
<td>.171*</td>
</tr>
<tr>
<td>7. Chat</td>
<td>.398**</td>
<td>.293**</td>
</tr>
<tr>
<td>8. Music listening</td>
<td></td>
<td>.221**</td>
</tr>
</tbody>
</table>

*Correlation is significant at the 0.05-level (2-tailed).
**Correlation is significant at the 0.01-level (2-tailed).
While listening to music may distract the students somewhat, downloading film or music is even more passive. It is done in the background while using the laptop for other purposes and does not include the actual use or consumption of the downloaded material. Nevertheless downloading was considered as intrusive and was clearly against the rules. Downloading was forbidden as it was considered as illegal; the students needed illegal software (file sharing software) to be able to download the files. Almost no students engaged in downloading during class; on average, less than 5% of the students did this on a daily basis.

The correlation between these two categories of use and the other use-categories is unreliable and there is no noticeable correlation pattern, to some extent because the amount of use (80% of the students listened to music every day, while 5% downloaded files) is so different.

**Mean values of laptop use**

To be able to discriminate possible differences between the participating schools and differences between genders, we calculated mean values and standard deviation for both sanctioned and unsanctioned use. The alternatives in the survey (Many times every day/ Occasionally every day - Occasionally every week – Occasionally every month – Never/ almost never) were given a value between 4-0, where the lowest value represents never engaging in the activity and the highest number represents daily use. The unsanctioned uses are web surfing, playing games and chatting.

*Table 4. Average and SD for unsanctioned use in 9th grade, divided by schools and gender*

Survey alternatives were assigned values between 4 – 0 (4 representing daily use and 0 representing never engaging in the activity).

<table>
<thead>
<tr>
<th>Average and SD</th>
<th>μ Web surfing (σ)</th>
<th>μ Playing games (σ)</th>
<th>μ Chat (σ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Girls</td>
<td>3.23 (1.46)</td>
<td>1.61 (0.97)</td>
<td>3.06 (1.61)</td>
</tr>
<tr>
<td>Boys</td>
<td>3.37 (1.46)</td>
<td>3.05 (1.53)</td>
<td>3.25 (1.57)</td>
</tr>
<tr>
<td>School 1</td>
<td>3.19 (1.44)</td>
<td>2.38 (1.55)</td>
<td>3.12 (1.59)</td>
</tr>
<tr>
<td>School 2</td>
<td>3.44 (1.47)</td>
<td>2.38 (1.41)</td>
<td>3.20 (1.60)</td>
</tr>
<tr>
<td>Total</td>
<td>3.31 (1.45)</td>
<td>2.38 (1.48)</td>
<td>3.16 (1.59)</td>
</tr>
</tbody>
</table>

We see only marginal differences between the two schools, as well as marginal differences between boys and girls concerning their unsanctioned use. However, as mentioned above in Figure 5, there is a significant difference between boys and girls concerning their engagement in playing games (highlighted in Table 4).

*Table 5. Mean values and SD for sanctioned use in the 9th grade, divided by schools and gender*

<table>
<thead>
<tr>
<th>Average and (SD)</th>
<th>μ Internet search(σ)</th>
<th>μ Word processing(σ)</th>
<th>μ Prepare presentations(σ)</th>
<th>μ Digital recording(σ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Girls</td>
<td>4.28 (0.83)</td>
<td>4.47 (0.73)</td>
<td>2.80 (0.98)</td>
<td>1.94 (0.97)</td>
</tr>
<tr>
<td>Boys</td>
<td>4.45 (0.83)</td>
<td>4.36 (0.90)</td>
<td>2.84 (1.04)</td>
<td>2.14 (1.07)</td>
</tr>
<tr>
<td>School 1</td>
<td>4.42 (0.76)</td>
<td>4.52 (0.73)</td>
<td>2.82 (1.08)</td>
<td>2.11 (1.0)</td>
</tr>
<tr>
<td>School 2</td>
<td>4.31 (0.91)</td>
<td>4.28 (0.92)</td>
<td>2.81 (0.92)</td>
<td>1.97 (1.05)</td>
</tr>
<tr>
<td>Total</td>
<td>4.37 (0.83)</td>
<td>4.41 (0.83)</td>
<td>2.82 (1.01)</td>
<td>2.04 (1.03)</td>
</tr>
</tbody>
</table>

Similarly, as described in Table 6, there are very small differences in sanctioned use between schools as well as in differences due to gender.

**Correlations**

In order to explore the relationships between the different types of use, both sanctioned and unsanctioned, correlation analyses were performed.

Table 6 shows that there is no or very low correlation between sanctioned and unsanctioned use. This means that it does not matter if the sanctioned use of the laptop increases – the students still use the laptops for unsanctioned
activities to the same extent. This refutes the teachers’ preconceived perception of the reciprocal relation between sanctioned and unsanctioned laptop use. Data show a significant \( p < 0.005 \) correlation (between 0.239 and 0.611) between all sanctioned uses. There is also a significant \( p < 0.005 \) correlation (between 0.421 and 0.505) between all unsanctioned uses.

### Table 6. Correlations between different types of use in 9th grade

<table>
<thead>
<tr>
<th>PEARSON CORRELATION</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Information search</td>
<td>.611*</td>
<td>.327*</td>
<td>.239*</td>
<td>.129</td>
<td>-.004</td>
<td>.044</td>
<td></td>
</tr>
<tr>
<td>2. Word processing</td>
<td>.293*</td>
<td>.247*</td>
<td>.002</td>
<td>-.032</td>
<td>.084</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Presentations</td>
<td>.306*</td>
<td>.052</td>
<td>.089</td>
<td>.142</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Digital recording</td>
<td>.206*</td>
<td>.001</td>
<td>.239*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Playing games</td>
<td>.444*</td>
<td>.421*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Web surfing</td>
<td>.505*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Chat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Correlation is significant at the 0.01 level (2-tailed).

**Correlation is significant at the 0.05 level (2-tailed).

**Discussion**

We have provided empirical evidence that there is not a reciprocal relationship between sanctioned and unsanctioned use of laptops in a 1:1 classroom and we cannot find a correlation between sanctioned and unsanctioned use. Hence, the teachers’ assumption that those students who are playing games and web surfing in a 1:1 classroom are the ones that are using the laptop less for schoolwork can be questioned.

Many activities in a 1:1-classroom, as well as in any classroom, may be disruptive and considered as counter-productive to student learning. Students do not always do exactly what is expected, and they do not always follow the teacher’s task design exactly. Laptops connected to Internet in the classroom provide students with opportunities to develop and elaborate tasks. Thus, student behaviour is not only a question of rules and regulations enforced by teachers, but also a matter of collectively developing an educational practice in which laptops are integrated. The design of a task in a 1:1 classroom is complex and must take a range of various components into consideration, all depending on the specific context. There is a need for reformulating and re-designing traditional tasks as a consequence of technical devices being accessible in the classroom all the time and more or less taken for granted. The teachers have to reflect on the students’ abilities, the time constraints, the learning goals, the demands of the curriculum, and the technical issues. Every situation in the classroom is unique and to solve the problems the teachers have to integrate pedagogy, content, and technology. To merely regard them as separate components is a “real disservice to good teaching” (Koehler & Mishra, 2008, p. 25).

The study shows that students’ daily, unsanctioned use was at the same level throughout the three years. However, as the survey only measured the frequency and not how much time that was spent, we cannot say how much of the students’ time was used on a particular activity, for example, playing games.

The rather high level of engagement in unsanctioned use also raises a question about the teachers’ reactions. What is defining the teachers’ opinions about whether an activity is considered unsanctioned or not? If a student, for example, listens to music using personal earphones while working on an assignment, this is usually considered as a sanctioned activity, but still not within the task. However, it is seldom considered as obtrusive or iniquitous as long as it does not have any disturbing effect on other students. Neither are activities, such as playing games or watching Facebook sanctioned. However, teachers apprehend these kinds of use differently to listening to music, mostly because they are considered as disruptive and provoking, and hence, they are often banned or limited. The teachers’ common reaction is that, from their point of view, the unsanctioned use of the laptop is similar to other forms of misconduct. As long as the behaviour does not disturb classmates, it seems that the use is accepted. Furthermore, it is apparent that students gradually elaborate a pattern of conduct that is adapted to the current circumstances. Most of them distinguish between when it is accepted or not accepted to take a micro pause, and a kind of mutual implicit agreement is developed between teachers and students. The question of the extent to which the unsanctioned and passive use has negative consequences for the individual student is still to be answered.
The article deals with the rarely explored routine use of technology in today’s classroom and contributes to the ongoing dialogue about how the classroom norms change over time. The results indicate the need to go beyond looking at unsanctioned use in isolation, and beyond considering it only as a disturbance, and rather to consider unsanctioned use in terms of the circumstances and the relational processes in the classroom.

The results in this study are partly dependent on self-reported data from the students, which can be considered as problematic. There is also a possible bias, because the students were the first in the municipality to get personal laptops, and some students reported that they had a feeling of being responsible for the successful implementation of the 1:1 project. Hence, they did not want to jeopardize the project by being unenthusiastic. The discussion concerning self-reported data and common method bias is well known (Conway & Lance, 2010). However, the usual assumption that common method bias inflates relationships between variables measured by self-report is questioned. For example, Spector (2006) reported that correlations among self-reported variables are near zero. Because we wanted to perform the study in an actual educational setting over an extended period of time, there were no possibilities of using control groups or an experimental setup. Evidence of construct validity in this article is provided through the use of mixed methods (Podsakoff, MacKenzie, Jeong-Yeon, & Podsakoff, 2003). The observations, interviews, and surveys paint an unequivocal picture of the laptop use. Due to an agreement made with the parents, it was not possible to discriminate the students on an individual basis. Hence, we could not follow changes on an individual level, which could be an issue for further research.

Conclusions

Two research questions were formulated. First, what is the relation between sanctioned and unsanctioned use in a 1:1 classroom, where student-use is regulated by discussing norms for use, rather than by introducing formal rules or technical restrictions? As has been shown in the results, there is no reciprocal correlation between sanctioned and unsanctioned use of the laptop. Students tend to use the laptop more for both sanctioned and unsanctioned use as the project continues. The results revealed that the percentage of students that never chatted or played games during class was increasing and the percentages of students that chatted or played games daily were stable at the same level during the three years. As has been shown in figure 4 and in figure 6 the percentage of students who never chatted or played games during class increased from 11% to 26% and 23% to 45% respectively over the three years. On the other hand, the percentages of students who chatted or played games daily remained at a similar level during the three years, between 49% and 52%, and 26% to 28% respectively.

The second question concerned how the use of laptops changes over time, for both sanctioned and unsanctioned uses. At the outset of our research, we expected to see how students, over time, would learn how to use their laptops in their everyday educational activities in an increasingly meaningful way, and that this would mean reduced activity in playing games, chatting and similar use. What we found was rather the opposite. When students increase their use of laptops and hence increase their skill, their use of laptops extends from use promoted by the teachers to a clearly unsanctioned or even forbidden use. Increased competence and longer experience seem to correlate positively with the activities in all types of use.

Finally, the issue concerning rules and regulations surrounding laptop use in classrooms is highlighted in this article. The students in the study did not have strict rules and the teachers were obliged to have ethical discussions about the use rather than forbid or restrain the laptop use. The results show that the unsanctioned use was relatively high, but on the other hand, it was stable and did not increase throughout the years. In addition, the number of students who did not game or chat at all increased every year. Further research is needed to investigate the relationship between rules and students’ use of laptops in the classroom.

References


The Effectiveness of Adopting E-Readers to Facilitate EFL Students’ Process-Based Academic Writing

Hui-Chun Hung¹ and Shelley Shwu-Ching Young²*

¹Institute of Information Systems and Applications, National Tsing Hua University, Taiwan // ²Institute of Learning Sciences, National Tsing Hua University, Taiwan // jimgo929@gmail.com // scy@mx.nthu.edu.tw

*Corresponding author

(Submitted August 28, 2013; Revised November 8, 2013; Accepted February 6, 2014)

ABSTRACT

English as Foreign Language (EFL) students face additional difficulties for academic writing largely due to their level of language competency. An appropriate structural process of writing can help students develop their academic writing skills. This study explored the use of the e-readers to facilitate EFL students’ process-based academic writing. The experiment was conducted in the graduate level class entitled “Technical and Scientific English Writing” in a northern Taiwan university for the entire semester, about 5 months. Students’ perceptions, writing outcome and portfolio were collected and were later evaluated. Both qualitative and quantitative approaches were adopted in this study. Empirical data were collected and analyzed to report on the performance of EFL students’ academic writing with and without the e-readers. Findings indicate that e-readers affected the process of reading, annotation, and information retrieval with the unique functions. For EFL students’ academic writing, e-reader can be a tool for reciprocal peer review that aided academic writing. Moreover, e-readers are significantly beneficial for students’ academic writing progress compared to the conventional paper-based materials. The functions of e-readers can assist students’ writing process and make the recursive circle of steps more efficiently. E-readers could afford creating a better writing environment in the process-based writing approach. This study further discussed the role of e-readers in the academic writing classroom and further discusses how to use e-readers to facilitate academic writing in the classroom.

Keywords

E-reader, Academic writing, English as Foreign Language (EFL), Process-based writing

Introduction

The English language has been increasingly highlighted as an important medium of communication in academia between English native speakers and non-native speakers (Leki, 2001; Zhu, 2004). However, academic writing is different for these two populations (Silva, 1993). English as Foreign Language (EFL) students face additional difficulties and stress for academic writing largely due to their level of language competency (Al Fadda, 2012; Bacha, 2002; Olivas & Li, 2006). To deal with the extra cognitive burden that EFL students are likely to experience for academic writing, an appropriate structural process of writing can help students develop their academic writing skills (Bacha, 2002). The writing process approach is a non-linear activity in which students need to go recursively through steps of planning, drafting, revising, editing, and publishing. It emphasizes on the continual interactions with the instructor and peers during the writing process (Greene, 2000; Tribble, 2002). Several studies have proposed a process-based writing instruction and the need for a supplementary tool for EFL students has been justified (Arslan & Şahin-Kızıl, 2010; Shang, 2007; Wang, Shang, & Briody, 2013). Meanwhile, the advancements of technology have led to the use of electronic reading systems for digital contents (Wright, Fugett, & Caputa, 2013). An electronic device such as the e-reader has the potential for aiding students in the process of writing. E-readers are generally portable and contain a built-in dictionary, and also tools for annotation and information browsing. However, few studies have examined the effect of adopting e-readers as an aiding device in academic writing for EFL students.

The aim of this study was to investigate the role of e-readers and how e-readers facilitated academic writing of EFL students. The following research questions guided this study:

• Can e-readers make any differences to EFL students’ writing performance, including writing outcome and portfolio, with or without the adoption of e-readers?
• What are the instructor and students’ perceptions on e-readers in the academic writing classroom?
• Can e-readers facilitate EFL students’ academic writing in a process-based approach? If yes, how do students use e-readers during the academic writing process?
• What are the roles of e-readers in the academic writing classroom?
Literature

Process-based writing and peer assessment

The process-based approach to writing or process writing has gained considerable attention and support from educators (Bacha, 2002). During process writing, writers are asked to repeatedly revise the drafts and this recursiveness is the main characteristic of the approach (Murray, 1978; Perl, 1980; Li, 1992). Given the difficulties of academic writing for EFL students, they need to have an appropriate structural approach in academic writing. The writing process used in this study was composed of six steps: (1) Prewriting, (2) Outlining, (3) Drafting, (4) Revising, (5) Editing and (6) Publishing (Otlowski, 1998).

Among the resources that writers can obtain during the writing process, feedback from teacher and peers is central, especially in a second language or foreign language setting because it promotes the sense of an audience in the students and sensitizes them to the needs of readers (Arslan & Şahin-Kızıl, 2010; Liu & Sadler, 2003). A number of studies have demonstrated the benefits of peer assessment on developing writing skills. Peer assessments are often conducted in peer response groups or through peer editing (Ferris, Brown, Liu, & Stine, 2011; Leki, 2001). It is also used with EFL classes (Liu & Hansen, 2002; Paulus, 1999; Villamil & de Guerrero, 1998). Peer assessment can be carried out when in the developing stage rather than at the end. It can also be conducted more frequently and immediately than teacher assessment. By interacting with peers and assessing the works of others, students themselves can also improve their self-assessment skills (Towler & Broadfoot, 1992). Although peer assessment can be less reliable and valid than that of the teacher’s assessment, especially when the assessors are young or inexperienced learners, peer feedback has the advantage of being available in greater volume and with greater immediacy than teacher feedback. Therefore, to implement process writing successfully, teacher’s supervision is essential to increase the reliability and validity of peer assessment (Topping & Ehly, 2001) and both the teacher feedback and peer feedback need to be integrated into the writing course (DiGiovanni & Nagaswami, 2001; Arslan & Şahin-Kızıl, 2010).

Technologies supporting students’ writing development

The advancement of the technologies has increased the opportunities of writing and made editing easier for students. More and more students engaging in creating and sharing digital content on the Internet have resulted in the participatory culture which shift the focus of literacy from individual expression to community involvement (Clinton, Purushotma, Robison, & Weigel, 2009). In the participatory culture, collective intelligence is one of the new literacies involved social skills through collaboration and networking (Clinton et al., 2009). Schäfer (2011) further defined the explicit participation which focused on active engagement of users in creative processes. The success of web technologies thrives on participation culture. An example is the online writing platform, the interactive features of Course/learning management system (CMS). The CMS provides a place where a student can share his writing with all the other students and the teacher can review and rate students’ works. The features that students indicated contributed the most to their learning, including sharing materials with peer learners, teacher’s feedback on assignments and online readings (Kvavik, 2005). Furthermore, studies indicate that students consider handheld devices useful and effective for reading and writing activities (Paredes, Sanchez-Villalon, Ortega, & Velazquez-Iturbide, 2007; Samuels, 2005).

Van Lehn, Chi, Baggett, & Murray (1995) suggested that peer assessment demands cognitive activities such as reviewing, summarising, clarifying, giving feedback, diagnosing errors and identifying missing knowledge or deviations from the ideal. In peer assessment, students have more opportunities to view assignments of peers. Kwok & Ma (1999) used the group support system to support collaborative assessment and found that online collaborative annotations can improve both learning achievement and motivation. However, previous works have mainly focused on collaboratively annotating electronic materials or documents (Lin & Lai, 2013; Su, Yang, Hwang, & Zhang, 2010; Yang, Chen, & Shao, 2004; Yeh & Lo, 2009). Adopting e-readers to facilitate students’ collaborative annotations in the process-based academic writing is scarce.
E-readers in the classroom

Various studies have examined the use of mobile technology to facilitate indoor or outdoor learning to provide learners and educators with support that is more active and adequate (Cavus & Ibrahim, 2009; Chen, Kao, & Sheu, 2003; Hung, Hwang, Lin, Wu, & Su, 2013; Wang, Young, & Jang, 2013). Previous studies have shown that mobile devices, such as PDAs, Tablets or smartphones, can be used as a cognitive tool to efficiently provide information and feedback relevant to the current learning situation (Hung & Young, 2013, Hwang, Chu, Lin, & Tsai, 2011; Vogel, Spikol, Kurti, & Milrad, 2010). Mobile learning has been recognized as being an effective learning approach (Chu, Hwang, Tsai, & Tseng, 2010; Rogers & Price, 2009). However, few studies have discussed the adoption of e-readers in the high education.

E-readers are portable electronic devices designed for reading digital books and electronic documents, such as the Amazon Kindle, the Sony Book Reader and the B&N NOOK. Devices such as e-readers and portable e-book reading devices have been introduced into the market with increasing growth of sales and have brought a considerable amount of research issues (Wilson, 2003, Pappas, 2009; Harris, 2010). Several researchers (Dearley & McKnight, 2001; Schoolnik, 2002; Thayer et al., 2011; Waller, 1986; Wilson, 2003) have developed e-reader prototypes and assessment methodologies to resolve the usability and interface issues in language classrooms. Many studies on e-book have focused primarily on reading digital content on a computer (Chang, 2005; Chao, Chen, & Chang, 2010; Chen, Guimbretiere, Dixon, Lewis, & Agrawala, 2008; Chou, 2012; Huang, 2013; Lin, 2009; Oakley & Jay, 2008; Sun, 2003). Borgman (2007) argues that the continued use of printed books and journals contributed to the slow adoption of E-readers in academia than in the general public. Moreover, studies indicated that it is time-consuming to convert the current learning materials to electronic forms for e-readers (Marmarelli & Ringle, 2009; Rae, 2011; Young & Lin, 2012).

Wright et al. (2013) compared the vocabulary and reading comprehension of students who used e-readers and paper-based books. It was found that although students do not differ in their language comprehension, they are more likely to search for extra information when engaged with digital text. Nevertheless, several pilot studies integrated e-readers in an academic setting in BOTH UK and US found that the checking, bookmarking, highlighting, and annotation functions were in high demand for the students (Behler, 2009; Simon, 2001; The Trustees of Princeton University, 2010; Wilson, 2003). However, it is still unclear whether or how e-readers can be integrated into classrooms to become a useful learning device for EFL learners. It is one of the issues worthy of our investigation.

Method

To investigates the effects of using e-readers in an EFL process-based academic writing class, this study was conducted in a graduate writing class entitled “Technical and Scientific English Writing” in a northern Taiwan university. This was a 5-month class, running for an entire semester. Students’ perceptions, writing outcome and portfolio were collected and were later evaluated. Both qualitative and quantitative approaches were adopted in this study. Empirical data were collected and analyzed to report on the performance of EFL students’ academic writing with and without the e-reader. Triangulation was used to improve reliability of the study.

Participants

“Technical and Scientific English Writing” course was lectured for the academic purpose writing. Based on the heavy burden of the teacher need to correct and comment on every EFL students’ writing, this course is usually limited to about 20-25 students. This study was conducted in a natural setting. Twenty-three graduate students enrolled in this course and were divided into the experimental group which used e-readers (EG, n = 12) and the control group which used the conventional textbooks (CG, n = 11) to learn the same content. These graduate students with diverse backgrounds enrolled in this course for the purpose of improving their academic writing and to publish their papers as journal articles. The students had achieved an intermediate level of proficiency because they had studied English for at least 7 years. None of them reported having prior experiences of using e-readers. The instructor had been teaching the course for five years and wanted to take this opportunity to explore the potentials of e-readers and to ease the burden in teaching and reviewing students’ writing.
About the e-readers

The e-readers used in this study were manufactured by a company in Taiwan which has been well-known for manufacturing green products to provide innovative, clean and energy-efficient solutions for environment conservation. The e-reader is a 13.1-inch color touch display, with low power consumption, wireless connectivity, USB 2.0 slot, memory card slot, handwriting recognition function and weighs about 750g. The e-readers provide the functions to assist students’ listening and note-taking in lecture, textbook reading and information searching.

Setting

During the study, students in the EG used e-readers as their personal learning devices both in the class and after school (Figure 1). Learning materials were transformed to a digital version to be used on the e-readers. Moreover, the Learning Management System (LMS) was implemented for students to exchange their annotation and review comments through the e-readers. Students in the CG received identical contents in paper prints. In this course, students needed to complete the write-up of a research paper, which includes the introduction, method and results sections. The six steps of the process-based writing (Otlowski, 1998), including prewriting, outlining, drafting, revising, editing, and publishing steps were carried out in this course for all students. To implement the writing process successfully, both the teacher feedback and peer feedback need to be integrated into the writing course. Students were asked to form small groups of three or four to assess each other. Peer assessment was conducted recursively among the drafting, revising and editing steps (Figure 2). Moreover, students had to hand in a portfolio which is a reflection of their writing progress every four weeks. At the end of the course, each student was required to submit their academic paper as the final report.

Data collection and analysis

Both quantitative and qualitative data were collected in this study. Students’ learning performance was evaluated based on their writing outcome and writing portfolio. The writing outcome included the grades on four writing assignments – the introduction, method, results and a complete research paper. The writing portfolio was a self-evaluation report on writing progress that each student had to hand in every four weeks. The instructor would grade each student’s writing outcome and portfolio. To understand the writing performance, this study collected the students’ grades for assignments and reports. A questionnaire was administered at the end of the semester to explore
students’ perceptions on the usability of e-readers and the effects on learning. The questionnaire consisted of 46 questions on a five-point Likert scale. The first part of the questionnaire contained 26 questions on system usability, which were adopted from Lund’s USE Questionnaire (Lund, 2001) and translated into Chinese by the researchers. The second part of the questionnaire were 20 questions on perceptions of e-readers’ functionality devised by the researchers, also on a five-point Likert scale. The third part of the questionnaire consisted of open-ended questions on the advantages and disadvantages of e-readers, how students used e-readers to enhance their academic writing and their suggestions if any. In addition, both individual and focus group interviews were conducted.

The scores of writing outcome, portfolio and questionnaire were computed and analyzed with the SPSS 12.0 statistical software package, and the related descriptive statistics were also computed. Independent-samples t test was used to detect significant difference among variables. To assess the validity of the questionnaire, this study determined the associations between the questionnaire scores by factor analysis. The results of factor analysis ($\chi^2 = 5148.462$, $df = 990$, $p = 0.000(0.001)$) indicates the construct validity of the questionnaire was satisfactory. Moreover, the internal consistency and reliability were tested by means of the Cronbach’s alpha coefficient, and the result for the sample as a whole was 0.94 (for the various domains, the coefficient ranged from 0.85 to 0.91), indicating that the questionnaire was acceptable with adequate internal consistency and reliability.

Interviewing is one of the most powerful tools to understand our fellow humans (Fontana & Frey, 2008). Therefore, this research utilized a qualitative approach that focused on understanding of the participants’ feelings, experiences and perceptions in relation to using e-readers to assist writing. All participants were interviewed at the middle and the end of the semester. Moreover, students’ activities during the class were video recorded and observations were made to participants’ behavior of using e-readers to assist writing. The data was analyzed thematically following a qualitative approach (Vaughan, Schumm, & Sinagub, 1996). The observations and interviews were transcribed verbally and then coded and reduced by searching specifically themes, including peer interaction, interaction with e-readers, the reflections and attitude toward using e-readers to assist writing in class. Further data was gathered by recorded interviews with four individual teachers, from two partnerships, one in the primary sector and one in the secondary sector. This data was also analyzed thematically using the same technique.

Results and findings

Learning performance of experiment group (EG) and control group (CG)

To respond to the first question, students’ learning performance was measured by grades on writing outcome and writing portfolio. Independent T-test was used to examine the effects of the two teaching approaches (with or without e-readers) in terms of writing outcome and writing portfolio.

<table>
<thead>
<tr>
<th>Source</th>
<th>Mean</th>
<th>SD</th>
<th>DF</th>
<th>t value</th>
<th>F value</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Writing outcome (four assignments) EG (n = 12)</td>
<td>80.06</td>
<td>5.56</td>
<td>21</td>
<td>.697</td>
<td>1.678</td>
<td>.209</td>
</tr>
<tr>
<td>Writing outcome (four assignments) CG (n = 11)</td>
<td>77.16</td>
<td>13.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* $p < .05$. EG: experiment group; CG: control group.

To examine the writing outcome, the independent T-test was applied to analyze the grades of three assignments and the term paper. Results in Table 1 indicate that the learning achievement of students in EG is higher than CG. However, this difference was not statistically significant ($t = .697$, $p = .209 > .05$).

<table>
<thead>
<tr>
<th>Source</th>
<th>Mean</th>
<th>SD</th>
<th>DF</th>
<th>t value</th>
<th>F value</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Writing Portfolio (four self-reports) EG (n = 12)</td>
<td>89.50</td>
<td>2.37</td>
<td>21</td>
<td>.788*</td>
<td>5.299</td>
<td>.032</td>
</tr>
<tr>
<td>Writing Portfolio (four self-reports) CG (n = 11)</td>
<td>87.95</td>
<td>6.38</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note:* $p < .05$. EG: experiment group; CG: control group.

With regard to the learning process, the independent T-test was applied to analyze average grades of students’ portfolios. Results in Table 2 indicate that the score of EG is higher than CG and this difference is statistically significant ($t = .788$, $p = .032 < .05$).
It can be said that although there was no difference on writing outcome between the two groups, students who adopted e-readers showed stronger engagement compared to students who were given printed materials. The adoption of e-readers was beneficial to the students’ writing process.

**Students’ perceptions of the role of e-readers in the academic writing classroom**

Following the second research question, the researchers analyzed the questionnaire and students’ responses during the interview to explore students’ perceptions of the role of e-readers in an academic writing classroom. As shown in Table 3, students’ perceptions with regard to e-readers were positive. The usefulness, ease of learning, and the functionality were rated above 4.20, and the ease of use and satisfaction were rated above 3.88. These results indicate that the students’ perception about the e-readers was positive. Moreover, the lecturer and students’ feedback during the interview also showed this tendency.

**Table 3. Students’ perceptions on e-readers in academic writing (n = 12)**

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usefulness</td>
<td>4.20</td>
<td>0.87</td>
</tr>
<tr>
<td>Ease of use</td>
<td>3.46</td>
<td>0.52</td>
</tr>
<tr>
<td>Ease of learning</td>
<td>4.25</td>
<td>0.97</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>3.88</td>
<td>0.76</td>
</tr>
<tr>
<td>Functionality</td>
<td>4.30</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*Note. From 5 to 1 points mean strongly agree, agree, neutral, disagree, and strongly disagree.*

During the interview, students emphasized that the e-reader was very easy to carry around, comfortable to read with, and easy to use (E1aFCh; E4aMWe; E1bFLi; E3bMEr; E3aMJe; E1cFAm). With regard to the advantages, students mentioned that e-readers are very convenient for learning (E1cFAm; E4cFLi; E2aFEI; E4bFGI) and many students found it enjoyable (E2aFEI; E4bFGI; E1bFLi; E4aMWe). Several students also expressed their interests in this innovative approach. One of the students (E2aFEI) stated that, “This is a totally different way to learn English. It’s so great to use this device (e-reader) in the classroom. I was able to negotiate with my group members and control the writing process.” Moreover, some encouraging findings from the students were found. Students explained that they feel the review and revision process is very efficient using the e-readers. Students said that, “The annotation function of the e-readers helps keep tracks of the comments from my classmates and based on the records I could revise my paper almost without omission (E1aFCh; E2bMAr; E4cFLi; E1cFAm).” Another student stated that, “Revising paper was very hard for me, but I still felt great when I contributed something to my classmates’ papers. I really prefer to use e-readers as an aid in the writing process (E2cFGI; E1aFCh; E2aFEI; E3aMJe; E4cFLi).” In spite of the advantages mentioned above, student also suggested that e-readers should provide more Chinese input methods (E4bFGI; E1cFAm). Besides, they were also concerned if the e-readers are easily damaged (E4aMWe; E2bMAr).

To acquire more feedback on the effect of adopting e-readers into the academic writing, this study interviewed the course instructor and two teaching assistants. The instructor was positive about the innovative approach. She stated that, “In the beginning, I was wondering if the use of e-readers would make any positive effect on students. However, I found that the students were active and positive about learning academic writing in the classroom.” The instructor also mentioned that e-readers were useful not only because they can involve a variety of learning materials and also, at the same time keep a record of writing process, such as the portfolio. The instructor expressed, “I think the most important function of the e-reader is that it integrates all the learning materials into just one device. Moreover, all reports, homework, and lecture notes for the whole semester can be saved in this device as a collection. The e-reader provides students with a medium to record and manage the evidence of their learning digitally such as their learning and writing portfolio.” The instructor further stated that e-readers are especially helpful for students’ reading comprehension. “E-reader is a helpful tool in assisting students with listening and note taking during the lectures. Moreover, it also provides students access to resources such as dictionaries, related papers, writing handbook and textbook while they write.”
Functions of e-readers facilitating EFL students’ process-based academic writing

The questionnaire and the interview with students indicated that they used the functions of e-readers to facilitate their writing very often during the writing process (Table 4). For each step of the academic writing process, students might use different functions. The functions that students rated most needed for their writing were annotation (free hand style), dictionary, web browsing and switching between files. These functions were deemed useful and beneficial to their academic writing. Moreover, this study interview with students to further understand how students use these function.

<table>
<thead>
<tr>
<th>Functionality of e-readers</th>
<th>Step of the process</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prewriting</td>
<td>Outlining</td>
</tr>
<tr>
<td>Internet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Web browsing</td>
<td>4.83</td>
<td>4.75</td>
</tr>
<tr>
<td>File transferring</td>
<td>3.83</td>
<td>3.83</td>
</tr>
<tr>
<td>Annotating Sharing</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Annotation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Underline</td>
<td>3.75</td>
<td>4.08</td>
</tr>
<tr>
<td>Highlight</td>
<td>4.08</td>
<td>3.89</td>
</tr>
<tr>
<td>Text</td>
<td>4.33</td>
<td>4.42</td>
</tr>
<tr>
<td>free hand</td>
<td>4.50</td>
<td>4.92</td>
</tr>
<tr>
<td>Others</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dictionary</td>
<td>4.42</td>
<td>4.50</td>
</tr>
<tr>
<td>Switching between files</td>
<td>4.75</td>
<td>4.33</td>
</tr>
<tr>
<td>key-words searching</td>
<td>4.33</td>
<td>4.08</td>
</tr>
</tbody>
</table>

Note. 1 = never; 2 = rarely; 3 = sometimes; 4 = often; 5 = very often.

Regarding the annotation function, students mentioned their favor. “There are several different ways of annotation in this e-reader and the free hand function is my favorite (E3aMJe).” “Using free hand function can draw my idea easily and quickly (E2cFCi).”

Regarding the dictionary function, students mentioned that “I need to use the dictionary very often to check the definition of the word (E1bFLi; E2bMAr; E4cFLi).” “Because English is not my native language, the dictionary in the e-reader is very useful for me (E2aFEl; E3aMJe).”

Regarding the web browsing function, students mentioned about “I always search the answer from the Internet, especially when I need to check the definition of a word or the grammar for a sentence (E1aFCh).” “It’s important that I can access the Internet during or after class, so I can share my note for each class (E1bFLi).” “It’s easy to download the lecture slides and handouts through the Internet so I can access the materials at any time instead of being limited to where my computer is (E1aFCh; E1bFLi; E3aMJe).”

Regarding the switching between files functions, students mentioned about “In the process of writing paper, I need to check the other references frequently; the design of switch reading between different files is very important (E1bFLi, E2aFEI).” “It is often necessary to switch from one to another. That can be done more easily using e-reader to meet my need. Moreover, the device can contain lots of article and easily to search (E4bFGi; E1cFAm).”

Regarding the other functions of e-readers, students mentioned about “It’s easy to carry and much lighter than the text book.” (E1bFLi). “The e-reader surprised me for its long battery life; (E2aFEI) I can use it for the course in the afternoon and continue using it for my homework during the evening without recharging.” “Its power is enough for the reading for a whole day.”
Role of e-readers in the steps of process-based academic writing

Qualitative data collected from interview and observation indicated that student utilized e-readers intensely and extensively in each step of the academic writing process to brainstorm, outline writing structure, compose and edit their drafts before publishing the final text (Table 5). As found in the interview data, participants illustrated their points by relating their own experiences during the six steps of the writing process.

In the prewriting step, students can identify the audience, purpose and related materials of this article via e-reader. They mentioned that “I like to search and collect the related information (E2bMAR; E4cFLi; E3aMJe).” “The web browsing tool in the e-reader is very useful for me to search and update the most recent academic materials. The articles from the Internet are displayed well on the screen of the e-reader (E1bFLi; E2bMAR).” “When I prepare the materials for my writing, I search the related content from the Internet via the e-readers. That is important for me to do that via the e-reader, because I focused on the content on the e-reader more intently than on the computer (E2aFEi).” Moreover, students mentioned that they use annotation function of e-reader to highlight the key points on the literature they searched. “Instead of printing on paper, I like to read related literature in the digital format on the e-reader for easily archiving and searching them later. (E1bFLi; E2bMAR; E3aMJe; E4cFLi)”

In the outlining step, students created the structure of the composition using the annotation tool of the e-readers. They mentioned that “When I barnstormed for the outline of the article, e-reader is the idea tool to writing down the ideas I generated (E1aFCh; E1bFLi; E3aMJe).” They preferred different types of annotations to outline their article. “I like to add notes via the virtual on-screen keyboard of the e-reader for outlining my article (E1aFCh; E1cFAm; E2aFEi; E3aMJe).” “I like to use free hand annotation function to drawing and writing the outline because it’s casual for me (E1bFLi; E2bMAR; E2cFCi; E4bFGi; E4cFLi).”

In the drafting step, students organized notes according to the outline on the e-readers and drafted their article with the word processor software on their personal computer. They mentioned that “When I compose the ideas according to the outline and start to draft my composition with my computer, the e-reader is the supplementary tool to show the outline and idea in previous steps (E1bFLi; E2cFCi; E3aMJe; E3bMEr; E4aMWe; E4cFLi).” Some of them further indicated that “Drafting my article on the computer and referring to the literature and outline on the e-reader is very convenient and efficient (E2cFCi; E4aMWe; E4cFLi).”

In the step of revising and editing, students shared their writing with peers through the Learning Management System and revised their articles follow comments and suggestions of peer and the instructor. Among the two steps and the previous drafting step, peer assessment was conducted recursively and students shared comments through e-readers. They mentioned that “By using E-readers, classmates and I can exchange the annotation and drafts more easily (E1aFCh; E1bFLi; E2cFCi; E3aMJe; E4cFLi).” “I can receive the feedback from peers immediately from the e-reader (E2cFCi; E4aMWe).” Students also indicate that “Follow the comment from the peers I can editing and revising my writing efficiently on the computer (E1aFCh; E2aFEi; E2cFCi; E3aMJe; E3bMEr; E4aMWe).”

<table>
<thead>
<tr>
<th>Step</th>
<th>Purpose &amp; Process</th>
<th>Adoption of e-readers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Prewriting</td>
<td>Identify the audience and purpose</td>
<td>Internet access</td>
</tr>
<tr>
<td></td>
<td>Brainstorm (within a group or on his/her own)</td>
<td>Annotation</td>
</tr>
<tr>
<td></td>
<td>Search related information</td>
<td>(highlight, underline)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>File upload to LMS</td>
</tr>
<tr>
<td>2. Outlining</td>
<td>Design layout</td>
<td>Annotation</td>
</tr>
<tr>
<td></td>
<td>Create the structure of the composition</td>
<td>(text, free hand)</td>
</tr>
<tr>
<td>3. Drafting</td>
<td>Organize notes and ideas according to the outline</td>
<td>Word processor (pc)</td>
</tr>
<tr>
<td></td>
<td>Compose the ideas according to the outline</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Follow the outline and draft</td>
<td></td>
</tr>
<tr>
<td>4. Revising</td>
<td>Revise basic structure</td>
<td>Annotation (text, free hand)</td>
</tr>
<tr>
<td></td>
<td>Add or delete sections</td>
<td>File upload to LMS</td>
</tr>
<tr>
<td></td>
<td>Share with peers (in the group) and the instructor</td>
<td>Sharing of annotated files with peers and instructors</td>
</tr>
<tr>
<td></td>
<td>Revise the content based on peers’ and the instructor’s comments</td>
<td></td>
</tr>
<tr>
<td>5. Editing</td>
<td>Check for logical theme development</td>
<td>Sharing of the e-reader to LMS</td>
</tr>
</tbody>
</table>

Table 5. Adoption of e-readers in process-based writing

257
Discussion

The differences of EFL students’ writing process with or without the adoption of e-readers

Although there was no difference in writing outcome between the two groups, students who adopted e-readers showed better performance in writing portfolio than those who were given printed materials. This study further examines why the adoption of e-readers was beneficial to the students’ writing process. The qualitative data collected from our interviews with students indicates that the supplementary use of e-readers makes students review, summarize, give feedback, and identify errors more easily, which are necessary in writing process (Van et al., 1995). Compared with the control group, the experiment group indicates that the e-readers are helpful for sharing materials and feedback with peers immediately. The results indicate that students consider handheld devices are useful and effective for reading and writing activities, which support relevant findings of the studies (Paredes et al., 2007; Samuels, 2005). For teachers, the peer feedback in the experiment group has the strength of being available in greater volume and immediacy than those in the control group. With the aid of e-readers, students can compare notes and comments through interpretations with others easily. The adoption of e-readers might lead students an active engagement in the writing processes resulting in the development of collective intelligence (Clinton et al., 2009) and thrive the explicit participation (Schäfer, 2011).

This study further discusses the web browsing and literature search function using e-readers, and how they affect academic writing process of EFL students. We want to explore how the functions of the learning devices could be appropriately used to facilitate academic writing. It was noticed that the information browsing, highlighting, and annotation functions were in high demand for the students in research writing, which further strengthen studies of integrated e-readers in academic settings in UK and US (Simon, 2001; Wilson, 2003; Behler, 2009; The Trustees of Princeton University, 2010). In addition, this study noticed that EFL students relied heavily on the Internet at each step of the process-based writing. For these graduate students, the Internet provides references and related literature, which can be used to help them outline their own writing, and online dictionaries provide language resources such as grammar and spelling that can be used for self-checking. The Internet accessibility of e-readers is often limited to e-book downloads in the current e-readers market. Considering graduate EFL students’ needs, we thus emphasize that the importance of the Internet accessibility is far greater than manufactures assumed, because it provides students with access to abundant research resources as powerful writing tools.

E-readers affording a better academic writing environment

E-readers have many advantages over printed textbooks because of the significant functions such as information browsing, key-words searching, annotating, and access to the Internet. Therefore, based on the results from the observations, interviews and questionnaire, we conclude that e-readers could afford creating a better writing environment in the process-based writing approach (Figure 3). This study further examines the potentials and the roles that e-readers play in facilitating EFL students in academic writing from the aspects of an individual, peers, and teachers:

For individual

E-readers are the personalized portable library that affords access to different kinds of references and materials. Moreover, e-readers also support the Internet access to search online dictionary, Wikipedia to check the grammar or vocabulary. In the information explosion age, there is a rising call for information management, especially for the EFL students who write technical articles. They need more immediate reference books and writing guidelines to support their academic writings. E-readers, a flexible and effective device to manage content, become more and more powerful and important.
Moreover, e-readers are the annotating tools with a rotatable paper-like screen. The vertical display provides a full-view of an academic paper as if it seems to be on an A4 paper, which makes reading on the screen as comfortable as on papers. Not only can the device be used to read but also to write on and annotate with highlights, underlines and different text colors. Thus saving and revising articles become more easily.

*Figure 3. E-readers affording a better academic writing environment*

**For peers**

The role of e-readers is the medium for sharing annotation and comments in peer assessment. Through the Internet, e-readers make exchanges of annotation and drafts more easily. With e-readers, students can receive abundant and immediate feedback from peers which is strongly correlated with effective writing. Therefore, students have more opportunities to view assignments of peers. In this study, students used e-readers to support collaborative assessment and annotations which both can improve their learning achievement through the writing process.

**For teachers**

E-readers are the convenient tools to develop and manage digital learning materials. With the Internet connection, teachers are more easily to organize and manage their digital teaching materials because of the reusability and adaptability of e-readers.

Moreover, e-readers are the aid for saving revised drafts and portfolios. Through the peer assessment with e-readers, teachers can obtain the immediate updated feedback from students. E-readers can keep the revised and annotated drafts and, therefore, reduce the burden of teachers when they have to evaluate and monitor the learning process of students.

**Conclusion**

This study is a long-term investigation of how e-readers could enhance the academic writing of EFL students. It discusses the pedagogical basis for the writing process, as well as the benefits e-readers can bring to the classroom. Findings indicate that e-readers affected the process of reading, annotation, and information retrieval with the unique functions. For EFL students’ academic writing, e-readers can be the tool for reciprocal peer assessment that aided academic writing. This study concludes with the role of e-readers in the academic writing classroom and further discusses how to use e-readers to facilitate academic writing in the classroom. With the research findings mentioned above, this study conclude that: (1) The e-reader is significantly beneficial for students’ academic writing progress and may potentially help students improve their writing compared to the conventional paper-based approach. (2) Both the students and instructor were positive on the usefulness, ease of use, ease of learning, satisfaction and
functionality dimensions of e-readers in the academic writing classroom. (3) This integration of the e-readers into the writing class was helpful throughout the writing process from an individual student, peer group and teacher’s perspective. E-readers have the potentials to assist the EFL students in academic writing and function as a handheld library, an annotating tool, a medium for sharing annotations and comments and a storage for revised drafts. Moreover, the functions of e-readers can assist students’ writing process and make the recursive circle of steps more efficiently. E-readers could assist creating a better writing environment in the process-based writing approach. In the meantime, more studies will be conducted in the future to explore the effects of e-readers in different domain knowledge.

**Limitation of this study**

This study was conducted in a natural setting in the graduate level course for exploring uses of e-readers in process-based academic writing. Due to the heavy burden of the teacher for correcting and commenting on every EFL student’s writings, the writing course is usually limited to about 20-25 students. It has to be admitted that this is relatively a small scale investigation. However, research findings supported by both the quantitative and qualitative data shed light on in-depth investigation of the roles of the e-readers in the academic writing. The findings in this study are significant and valuable but the results could not be generalized to the other subjects.

**Acknowledgements**

The authors would like to thank Delta Electronics and National Tsing Hua University for their co-sponsorships of this project as well as the partial sponsorship from National Science Council Taiwan under the project code of NSC 101-2511-S-007 -002 -MY3.

**References**


261


Waller, R. (1986). What electronic books will have to be better than. *Information Design Journal, 5*(1), 72-75.


Effects of Worked Examples Using Manipulatives on Fifth Graders’ Learning Performance and Attitude toward Mathematics

Chun-Yi Lee¹* and Ming-Jang Chen²

¹Center for Teacher Education, National Taipei University, Taiwan, No.151, University Rd., San-Shia Dist., New Taipei City 23741, Taiwan // ²Center of General Education, National Chiao-Tung University, Taiwan, No. 1001, University Rd. Hsinchu 30010, Taiwan // chunyi.lii@gmail.com // mjchen@mail.nctu.edu.tw

*Corresponding author

(Submitted August 26, 2013; Revised January 28, 2014; Accepted February 2, 2014)

ABSTRACT

The purpose of this study was to investigate the influence of worked examples using virtual manipulatives on the learning performance and attitudes of fifth grade students toward mathematics. The results showed that: (1) the utilization of non-routine examples could promote learning performance of equivalent fractions. (2) Learning with virtual manipulatives was as effective as with physical manipulatives. (3) Virtual manipulatives can increase learning enjoyment compared with physical counterparts. Implications and recommendations are also presented for future research.

Keywords

Non-routine example, Equivalent fraction, Worked example, Virtual manipulative

Introduction

The concept of equivalence plays an important role in learning fractions. Students must first be equipped with basic concepts related to fractions before they can learn the meaning of equivalent fractions and it is only after these two steps that the concept of rational numbers can be further developed. The equivalence relation of fractions must be taken into consideration when comparing the order relation between two fractions. An understanding of the concept of equivalent fractions is the basis for performing the four arithmetic operations (addition, subtraction, multiplication and division) on fractions with different denominators. Equivalent fractions are the most difficult of all sub-concepts related to fractions, requiring that students have flexibility in their thought processes and a willingness to solve problems by advancing from concrete operations into formal operations. However, among elementary school students, the comprehension of the concept of equivalent fractions tends to be quite weak (Kamii & Clark, 1995). Elementary school students in Taiwan also have this problem; in fact, many students are unable to fathom that one half is equal to two quarters, even if they have previously studied equivalent fractions (Yu & Leu, 2002). An incomplete understanding of fractions or overly rigid thinking can prevent students from solving problems related to equivalent fractions.

Most elementary school students are at the stage of concrete operations, in which action and iconic representation can only be formed through actual practice. Therefore, physical manipulatives are commonly used in mathematics education to make abstract ideas and symbols more meaningful and comprehensible to students (Durmus & Karakirk, 2006). Developing a comprehension of concepts related to equivalent fractions requires that objects be split into different portions; however, physical manipulatives are ill-suited to the arbitrary splitting of objects. Moyer, Bolyard, and Spikell (2002) recently defined virtual manipulatives as interactive, web-based visual representations of a dynamic object capable of facilitating the development of mathematical concepts by students. Virtual manipulatives allow the splitting of objects into different portions to facilitate the learning of concepts associated with equivalent fractions. This enables students to visualize specific, abstract mathematical concepts and the content of learning can be broken down for presentation purposes (Chang, Yuan, Lee, Chen, & Huang, 2013). This makes virtual manipulatives excellent tools to facilitate the learning of concepts associated with equivalent fractions.

Worked examples have been successfully applied in the instruction of computer programming, algebra, and geometry (Carroll, 1994; Paas & van Merrienboer, 1994). Students with experience using worked examples as an instructional strategy adopt problem-solving techniques more quickly and present heightened problem solving performance (Chandler & Sweller, 1991). The provision of appropriate worked examples can aid in the formation of concepts by learners. Virtual manipulatives and worked examples offer exciting new possibilities as learning aids; however, there’s a noticeable lack in the literature of the large-scale validation of these strategies and few studies have
explored the practical implementation of combining worked examples with virtual manipulatives. Understanding the properties of worked examples with virtual manipulatives and how these relate to learning is important in predicting which material will be more beneficial and can help to inform the design of new learning materials. This study explores how different approaches to example integration influence the learning performance of fifth grade students.

**Literature review**

**Factors affecting the learning of equivalent fractions**

A failure to learn the concept of equivalent fractions is due mainly to overly rigid thinking (Peng & Leu, 1998). Many factors influence the learning of equivalent fractions, such as the capacity for flexible thinking, combining, operational thinking, and unit forming, as outlined in the following.

In the graphic representation of continuous quantity, flexible thinking capability refers to the ability of learners to realize that fractions with different names are actually the same fraction (for example: 2/6 and 3/9 both equal 1/3) through visualization and ignorance of the split line. The ability to perform these procedures is largely determined by the number of split lines and whether the split regions are interconnected. Many students insist that they can only accept an equivalent name for a fraction when the denominator of the fraction is equal to the number of split blocks and when the split blocks are interconnected (Behr, Wachsmuth, Post, & Lesh, 1984; Behr & Post, 1992). Booth (1987) interviewed 11-year-old school students in Britain and found that 95% of them believed that Figure A in Figure 1 was 1/3, while 73% of them thought that Figure B was 1/3; a difference of 22%. The problem was that the students believed that 2/6 is different than 1/3. This is an indication that if the learner has a graphic image of the problem or learns to ignore the split line, another name can be generated for the equivalent fraction.

In the graphical presentation of discrete quantities, flexible thinking capability refers to a learner’s ability to utilize actual manipulatives or their imagination to conduct re-partition or re-combination in the process of problem solving (Behr et al., 1984). For example, when we use small circles to solve the problem of \( \frac{2}{3}=\frac{4}{6} \), the first thing to do is convert the figures into fractions. First, each pair of circles in Figure 2 is regarded as one set, such that six circles become three sets. Four of the six circles are black, so they can be converted into the fraction 4/6. Two of the three sets of circles are black, so they can be converted into the fraction 2/3. There are exactly the same numbers of black circles, leading to the result that 4/6=2/3.

![Figure 1](image1.png)  
*Figure 1. Illustration of flexible thinking with respect to equivalent fractions based on the scenario of continuous quantity*

![Figure 2](image2.png)  
*Figure 2. Illustration of flexible thinking with respect to equivalent fraction based on discrete quantity scenarios*

“Combination capability” refers to the students’ ability to use specific problem solving strategies to deal with problems of equivalent fractions or residual fractions. Students demonstrate this capability by dividing a unit quantity into several parts and applying the correct treatment to every part. Combination capability is the ability to combine every treated quantity into a designated fraction of unit quantity (Peng & Leu, 1998). “Operational thinking capability” refers to the ability to use different splitting approaches on the same figure, i.e., a situation in which there are equal areas, but different shapes (Kamii & Clark, 1995).

Knowledge can be classified into the figurative and operative aspects. The figurative aspect of knowledge is based on the observed shape of knowledge, while the operative aspect is based on relation knowledge, a process which cannot be observed. For example, there are different ways to split the rectangle into two halves, resulting in either rectangles...
or triangles. Using just the shape aspect of knowledge, the rectangles and triangles differ in shape and size. Using the relation aspect of knowledge, a student might understand the relationship between the split triangles/rectangles and the original rectangle. In this manner, the student can infer that the triangles and rectangles actually have the same size (1/2 of the original rectangle) without being visually misled. The key for the learner is whether or not they can find an appropriate unit within the figures to precisely split the designated part in order to solve problems of equivalent fractions. The reverse capability is the ability to use the unit to recombine the figure into a whole entity after the exact splitting of the appropriate units; this is called the unit forming capability (Saenz-Ludlow, 1994, 1995). For example, Saenz-Ludlow (1994, 1995) provided several b1, b2 and b3 figures (shown in Figure 3) to third-grade students. At first, the students were allowed to compare b1 with the whole entity and then compare b2 and b3, respectively. Using these three pieces, students were able to divide the whole entity (Figure 3), such that b2 is 1/3 of the whole entity. In this study, when children were first asked to compare b1 with b2, they could not find the correlation between them; however, after receiving a hint using b3, they quickly realized that b2 is 2/3 of b1. This indicates that the students knew that every small unit must be equal to each other after splitting, which enabled them to adjust the small units in order to come to a solution. When b3 was presented, the students said that b2 is 2/3 of the half piece (b1). B3 gave them a representation of the relationship between b1 and the whole entity, and 2/3 represents the relationship between b2 and b1.

![Figure 3. Illustration of unit forming with equivalent fractions](image)

**Current status of teaching materials for equivalent fractions**

As indicated by the mathematics curriculum standards of the Nine Year Integrated Curriculum Syllabus in Taiwan, basic fractions are introduced from first grade to third grade, while the meanings of equivalent fractions, fraction reduction/expansion, and the addition/subtraction/multiplication/division of fractions are introduced from fourth to fifth grades (Taiwan Ministry of Education, 2004).

In Taiwan, all versions of equivalent fraction teaching materials are presented without the direct use of terms such as fraction expansion and fraction reduction, and without the direct introduction of an operation approach wherein the value of a fraction remains unchanged, with both sides of the equations multiplied or divided by the same number. Instead, smaller unit components are generated through re-splitting or re-combination in specific scenarios, or larger unit components are generated through merging or re-combination, while the equivalent relationships of components represented by these two kinds of unit components are compared to each other. Under the scenario of continuous quantity, two strategies can be adopted: resorting to intuitive experience and resorting to the number of split parts. Resorting to intuitive experience refers to the approach of splitting ribbons or pies into different parts and comparing the lengths and sizes of different quantities to recognize the equivalence of two fractions in terms of length or area and concluding that the two fractions are equal. Resorting to the number of split parts refers to the direct notification of the number of split parts and determination of whether or not the two fractions are equal to each other with the approach of number line. Under scenarios of discrete quantity, the strategies of “resorting to the content” and “resorting to merger or re-combination of the content” are most commonly adopted for problem solving. Resorting to the content refers to the determination of equivalence through a comparison of quantity of contents. Resorting to merger or re-combination of the content refers to the approach of finding other names for the fraction by regarding multiple contents as a single entity and determining whether the quantities are identical.

Analyzing the content of the Taiwanese teaching materials for equivalent fractions led us to the conclusion that the existing examples addressed the issue of flexible thinking in continuous quantity and discrete quantity scenarios. The examples allowing flexible thinking in scenarios of continuous quantity include interconnected split blocks, while the examples based on scenarios of discrete quantity deal with activities involving the splitting or combining of groups of identical objects. Thus, providing worked examples of non-continuous colored blocks in the scenario of
continuous quantity, and demonstrating the existence of more than two kinds of objects in equivalent fraction problems in group mode, may be effective techniques with which to improve the flexible thinking ability of students.

**Virtual Manipulatives**

A virtual manipulative is very similar to a physical manipulative. It is a dynamic object with interactive features, which can be put on a website. The presentation of this kind of dynamic object provides students with an opportunity to construct mathematical knowledge (Moyer, Bolyard, & Spikell, 2002; Moyer, Niezgoda, & Stanley, 2005). Virtual manipulatives are equipped with the following features (Yuan, Lee, & Wang, 2010; Chang, Yuan, Lee, Chen, & Huang, 2013): (1) variability, in which the learner can color parts of the objects, or increase or reduce the quantity of certain object; (2) unlimited supply, which resolves the issue of an insufficient number of physical manipulatives in class; virtual manipulatives also save teachers from the time consuming distribution and organization of teaching aids, and makes it easy to arrange teaching aids simply by clicking the recycle bin icon to clear all teaching aids from the screen; (3) simultaneously presentation of figures and symbols on the screen to enhance the link in the learner’s mind between these two presentations of quantity.

Empirical evidence related to the use of virtual manipulatives for mathematics learning in the classroom is still relatively new and somewhat limited. Yuan, Lee, and Wang (2010) examined the effects of using virtual manipulatives and physical manipulatives in the study of polyominoes on junior high school students. In that study, the group using virtual manipulatives learned as effectively as the group using physical manipulatives. In addition, the use of virtual manipulatives enabled the generation of new ideas and helped to develop an appreciation of symmetry and the effectiveness of rotating figures. Manches, O’Malley; and Benford (2010) indicated that differences in the properties of manipulatives (comparing virtual and physical manipulatives) might influence the numerical strategies employed by students. Suh and Moyer (2007) examined the effects of developing representational fluency using virtual and physical manipulatives. The results showed that despite differences in manipulative models, both the physical and virtual environments are effective for learning and encouraging relational thinking and algebraic reasoning.

In contrast, Chang et al. (2013) adopted a non-equivalent quasi-experimental design, and recruited participants from two classes of third grade students in an elementary school in Taiwan. Their results demonstrated the effectiveness of virtual manipulatives over that of physical manipulatives on three subscales of immediate learning performance and all four subscales of retention performance. Remier and Moyer (2005) reported that third grade students using virtual manipulatives to learn fractions showed statistically significant gains in the development of conceptual knowledge. Student surveys and interviews have indicated that manipulatives capable of providing immediate and specific feedback are easier to use than traditional methods, and enhance enjoyment while learning. These studies demonstrated that virtual manipulatives offer unique advantages and can be as or more effective than physical manipulatives in the support of learning. However, it is more important to examine the characteristics of learning environments using manipulatives and how these characteristics influence learning experiences. Therefore, this study explored the effects of various worked examples using virtual or physical manipulatives on the learning of equivalent fractions.

**Method**

**Research design**

This study adopted a quasi-experimental design with the “method of example integration” as an independent variable. Based on differences between worked examples using virtual or physical manipulatives, example integration can be categorized as traditional continuous examples (TCE), technology supported continuous examples (TSCE), and technology supported mixed examples (TSME). The dependent variables were “the learning performance of equivalent fractions” and “mathematics attitudes.” The learning performance of equivalent fractions refers to the learning outcomes after the experiment: (1) basic flexible thinking: basic inpainting capability and splitting capability; (2) advanced flexible thinking: advanced inpainting capability, combination capability, operational thinking capability, and unit forming capability. Attitude toward mathematics learning refers to the
learners’ perception of mathematics following the experimental teaching, including learning enjoyment, learning motivation, and reported anxiety caused by studying mathematics.

Participants

The participants in this study included 5th grade students from an elementary school in Taipei City. All of the participants had learned basic concepts related to fractions, such as decimals, simple fractions, and unit quantity before this teaching experiment was conducted. As such, they were equipped with preliminary knowledge of equivalent fractions but without having learned the concept of equivalent fractions. In order to coordinate with the curriculum of the original class, we randomly selected 100 participants from three classes out of 13 classes in the 5th grade, and randomly assign 34 students for the TCE group, 32 students for the TSCE group, and 34 students for TSME group. The sample was 90 students after eliminating learners who failed participate fully. This left 30 students left in the TCE group, 30 students left in the TSCE group, and 30 students in the TSME group.

Instruments

Teaching materials

The main purpose of this study was to explore the effects of various approaches to example integration on the learning of equivalent fractions by students in the 5th grade. These approaches to example integration can be categorized as “continuous examples” and “mixed examples.” Continuous examples refer to continuous problems of equivalent fractions, while mixed examples refer to solving discontinuous and continuous problems of equivalent fractions. The various approaches to teaching were “traditional instruction” and “technology supported instruction.” Traditional instruction refers to the instructor explaining the concept of equivalent fraction in conjunction with physical manipulatives and providing students with opportunities to use manipulations to verify mathematical concepts. Technology supported instruction refers to the teacher explaining the concept of equivalent fractions in conjunction with virtual manipulatives and providing students with appropriate opportunities to use manipulations to verify the mathematical concepts. Virtual manipulatives in this study included a “Magic Board” and “Fraction Bar” developed by Yuan & Lee (2012). Magic Board maintains the properties of physical manipulatives and enables a clearer representation of mathematical concepts. This tool is available on the internet (http://163.21.193.5). Another virtual manipulative is the Fraction Bar, which is an interactive tool used to explore the concepts of fractions in length mode.

According to the approach adopted for example integration, the students were divided into TCE, TSCE, and TSME groups. The difference between the TCE and TSCE groups was the use of the different manipulatives, in which TCE group verified the concepts of equivalent fractions using physical manipulatives and the TSCE group verified concepts of equivalent fractions using virtual manipulatives. These two groups had identical examples and approaches to the explanation of concepts. The differences between the teaching materials used in TSCE and TSME are listed in the following:

- TSCE

In a continuous quantity scenario, continuous examples refer to colored blocks that are continuous and non-scattered, such that the learners can refer to the same fraction using different names through visualization and ignorance of the split line. Examples of continuous equivalent fractions include equivalent fraction problems in the length mode and area mode. Take for example the length mode, as shown in Figure 4. In order to identify a fraction that is equivalent to 1/2, the learner can use the virtual manipulative (fraction bar) to figure out the appropriate splitting approach and then apply color to find a colored ribbon that is equivalent to 1/2 of the colored ribbon. The fraction bar can only color the figure by dragging, which prevents the scattering of the colored blocks. We can use the alignment line to ensure that every fraction figure has the same length.

In discrete quantity scenarios, continuous examples refer to the equivalent fraction problem in group mode, in which re-splitting or re-combination is conducted using a group of identical objects to demonstrate that two fractions are
equal. A comparison between the quantity of $\frac{2}{5}$ a box of apples and $\frac{6}{15}$ a box of apples is as shown in Figure 5. Virtual manipulative can be used to reduce the number of split lines in order to identify the most appropriate splitting approach and to provide a coloring function to help students realize these two fractions are equal.

- **TSME**

In addition to providing learners with examples of continuous equivalent fractions, mixed examples can also provide non-routine examples that cannot be seen in traditional teaching materials. In the scenario of continuous quantity, non-routine examples refer to colored blocks that are not continuous, in which the learner must first arrange discontinuous blocks into continuous blocks and refer to the same fraction using different names through visualization and ignorance of the split line. Non-routine examples of equivalent fractions in the scenario of continuous quantity include problems of equivalent fractions in length mode and area mode. Take the non-routine example of the equivalent fraction of length mode, as shown in Figure 6. To obtain the answer for $\frac{1}{3} = ?/6$, the learner can use virtual manipulatives to add or eliminate split lines, identify an appropriate splitting approach, color any small frame of the fraction bar with six equal portions, move all discontinuous blocks together to form continuous blocks as a figure of the fraction, and arrange the fraction bar into three equal portions to verify that $\frac{1}{3}$ is equaled to $\frac{2}{6}$.

*Figure 4. Example of continuous equivalent fraction in length mode*

*Figure 5. Example of continuous equivalent fraction in group mode*

*Figure 6. Integrated example of equivalent fractions in length mode*

*Figure 7. Integrated example of equivalent fractions in group mode*
In discrete quantity scenarios, non-routine examples refer to situations in which there are more than two kinds of objects in the equivalent fraction problem in group mode. The learner must first re-arrange the objects to determine an appropriate splitting approach in order to realize that the two fractions are equal. In the example in Figure 7, finding the equivalent fraction of 16/24 involves scattered objects, such that the learner must first move the figures and re-arrange them into a single object. The coloring function is then used to identify an appropriate splitting approach for drawing split lines and regarding multiple objects as a group to understand that the two fractions are equal. In this manner, the students obtain an equivalent fraction.

**Equivalent Fraction Achievement Test**

The purpose of the equivalent fraction achievement test was to evaluate learning performance with respect to equivalent fractions after conducting teaching experiments involving different approaches to example integration. The test was meant to evaluate Taiwanese teaching materials related to equivalent fractions. The test items were used to measure two types of thinking, “basic flexible thinking” and “advanced flexible thinking.” Basic flexible thinking involves the basic thought processes involved in resolving equivalent fractions, divided into basic inpainting capability and splitting capability. Basic inpainting capability is a measure of the learner capacity associated with solving continuous equivalent fractions under scenarios of continuous quantity, used to determine whether the learner has the flexible thinking. Splitting capability measures students’ ability to solve continuous equivalent fractions in scenarios of discrete quantities, as a reflection of the thought processes involved in re-splitting or re-combining. Advanced flexible thinking involves solving problems associated with discontinuous equivalent fractions. This advanced form of flexible thinking includes four categories: advanced inpainting, combination, operational thinking, and unit forming. Advanced inpainting involves the solving of discontinuous equivalent fractions under scenarios of continuous quantity, regardless of whether the learner can assimilate discontinuous blocks or employs flexible thinking. Combination involves solving problems associated with discontinuous equivalent fractions in scenarios of discrete quantity, as a reflection of the thought processes required for the re-arrangement of more than two kinds of objects as well as re-splitting or re-combining. Operating thinking involves solving problems of equivalent fractions associated with objects of various shapes (i.e., the problem of equal areas within different shapes), and the inference that all figures are equal, despite visual clues to the contrary. Unit forming involves solving problems associated with the precise division of the whole entity using appropriate units, and using these units for recombination into a whole entity. This primary involves the identification of a unit suitable for the division of the object into designated parts.

The items in the achievement test for problems of equivalent fractions measure two types of thinking: basic flexible thinking and advanced flexible thinking. Basic flexible thinking type can be further divided into two categories and advanced flexible thinking can be further divided into four categories, resulting in a total of six forms of flexible thinking capabilities. The test included four problems for each of these capabilities, resulting in a total of 24 problems. One point was assigned for each problem (24 points in total). The reliability of these problems was qualified using a test of internal consistency. The values for Cronbach’s alpha were as follows: basic inpainting (α = 0.70), splitting (0.75), advanced inpainting (α = 0.65), combination (α = 0.83), operational thinking (α = 0.60), and unit forming (α = 0.84). The score for the entire test was α = 0.92, which is an acceptable coefficient of internal consistency. The difficulty of the problems was between 0.57 ~ 0.83, and the item discrimination index was between 0.33 ~ 0.86. These results demonstrate that the achievement test for problems of equivalent fractions has appropriate difficulty and discrimination index.

**Questionnaire of Attitudes toward Mathematics Learning**

The purpose of the Questionnaire of Attitudes toward Mathematics Learning was to reveal the feelings of learners related to learning equivalent fractions through various approaches to example integration. The content of the questionnaire covered three categories: learning enjoyment, learning motivation and anxiety caused by studying mathematics. A total of 15 problems (five problems for each dimension) were included to deal with attitudes toward mathematics learning. This included positive items for learning enjoyment and learning motivation, and negative items for the anxiety caused by studying mathematics. A five-point Likert Scale was adopted. Scoring for positive items involved awarding 1 point (strongly disagree) to five points (strongly agree); scoring for negative items was just the opposite. The reliability of this questionnaire was verified using an internal consistency test, the values of which are as follows: Cronbach’s alpha for learning enjoyment (α = 0.83), learning motivation (α = 0.85), anxiety
caused by studying mathematics ($\alpha = 0.72$), and for the entire questionnaire ($\alpha = .90$), indicating excellent internal consistency.

Results

Analysis of learning performance in equivalent fractions

Basic flexible thinking

This study employed multivariate analysis of covariance to explore the performance of basic flexible thinking in the solving of equivalent fractions, the results of which are presented in Table 1. The effects of example integration reached the level of significance ($\text{Wilks' Lambda} = .749, p < .001, \eta^2 = .134, \text{Cohen d} = .787$) indicating that as far as the performance of basic inpainting and splitting, a significant difference was observed in the average score of at least one capability in groups using different approaches to integration, as shown in Table 1. The influence of example integration method on basic inpainting capabilities was significant ($F_{(2,83)} = 10.106, p < .001$), and the influence of example integration method on splitting capabilities was non-significant ($F_{(2,83)} = 1.975, p = .145$), indicating that the basic inpainting capabilities of learners receiving TSME ($M = 3.513$) was significantly better than those receiving TCE ($M = 2.662$) and TSCE ($M = 2.825$), and no significant difference was observed in the splitting capabilities of learners using various approaches to example integration.

<table>
<thead>
<tr>
<th>Source of variance</th>
<th>Basic flexible thinking</th>
<th>$SS$ (Type III sum of squares)</th>
<th>$df$ (Degree of freedom)</th>
<th>$MS$ (Sum of squares)</th>
<th>$F$ (F test)</th>
<th>Sig. (Significant)</th>
<th>Post comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior knowledge</td>
<td>Basic inpainting capability</td>
<td>40.329</td>
<td>1</td>
<td>40.329</td>
<td>66.498*</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Splitting capability</td>
<td>46.953</td>
<td>1</td>
<td>46.953</td>
<td>46.776*</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>Example integration approach</td>
<td>Basic inpainting capability</td>
<td>12.258</td>
<td>2</td>
<td>6.129</td>
<td>10.106*</td>
<td>.000</td>
<td>TSME&gt;TCE</td>
</tr>
<tr>
<td></td>
<td>Splitting capability</td>
<td>3.965</td>
<td>2</td>
<td>1.983</td>
<td>1.975</td>
<td>.145</td>
<td>=TSCE</td>
</tr>
<tr>
<td>Error</td>
<td>Basic inpainting capability</td>
<td>50.337</td>
<td>83</td>
<td>.606</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Splitting capability</td>
<td>83.314</td>
<td>83</td>
<td>1.004</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$p < .05$.

Analysis on the advanced flexible thinking performance of equivalent fraction

This study employed multivariate analysis of covariance to explore the performance of advanced flexible thinking in solving equivalent fraction. The influence of the example integration method was significant ($\text{Wilks' Lambda} = .778, p = .009, \eta^2 = .118, \text{Cohen d} = .732$), indicating that there was a significant difference in the average score for at least one capability in each group of the sample. These differences demonstrate that learners using various approaches to example integration differed in their advanced flexible thinking required for inpainting, combination, operational thinking, and unit forming.

The advanced flexible thinking required to solve equivalent fractions includes advanced inpainting, combination, operational thinking, and unit forming. A summary of covariance analysis is presented in Table 2. The effects of advanced inpainting ($F_{(2,83)} = 9.096, p < .001$) and operational thinking ($F_{(2,83)} = 4.691, p = .012$) reached the level of
significance, indicating that these capabilities were stronger among learners receiving TSCE, compared to those receiving TCE or TSCE. No significant differences were observed between those receiving TCE and TSCE in the performance of advanced inpainting or operational thinking. The effects of the various approaches to integration on combination capability and unit forming capability were non-significant (combination capability: $F_{(2,83)} = 1.311$, $p = .275$; unit forming capability: $F_{(2,83)} = 2.137$, $p = .124$).

Table 2. Summary of covariance analysis on performance of advanced flexible thinking for equivalent fraction

<table>
<thead>
<tr>
<th>Source of variance</th>
<th>Advanced flexible thinking capability items</th>
<th>SS (Type III sum of squares)</th>
<th>df (Degree of freedom)</th>
<th>MS (Sum of squares)</th>
<th>$F$ (F test)</th>
<th>Sig. (Significant)</th>
<th>Post comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior knowledge</td>
<td>Advanced inpainting capability</td>
<td>30.703</td>
<td>1</td>
<td>30.703</td>
<td>30.692*</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Combination capability</td>
<td>70.566</td>
<td>1</td>
<td>70.566</td>
<td>89.327*</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Operational thinking capability</td>
<td>.016</td>
<td>1</td>
<td>.016</td>
<td>.017</td>
<td>.897</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unit forming capability</td>
<td>59.096</td>
<td>1</td>
<td>59.096</td>
<td>36.431*</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>Example integration approach</td>
<td>Advanced inpainting capability</td>
<td>18.199</td>
<td>2</td>
<td>9.100</td>
<td>9.096*</td>
<td>.000</td>
<td>TSME&gt;TC E=TSCE</td>
</tr>
<tr>
<td></td>
<td>Combination capability</td>
<td>2.072</td>
<td>2</td>
<td>1.036</td>
<td>1.311</td>
<td>.275</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Operational thinking capability</td>
<td>8.875</td>
<td>2</td>
<td>4.438</td>
<td>4.69*</td>
<td>.012</td>
<td>TSME&gt;TC E=TSCE</td>
</tr>
<tr>
<td></td>
<td>Unit forming capability</td>
<td>6.935</td>
<td>2</td>
<td>3.467</td>
<td>2.137</td>
<td>.124</td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>Advanced inpainting capability</td>
<td>83.030</td>
<td>83</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Combination capability</td>
<td>65.568</td>
<td>83</td>
<td>.790</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Operational thinking capability</td>
<td>78.517</td>
<td>83</td>
<td>.946</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unit forming capability</td>
<td>134.637</td>
<td>83</td>
<td>1.622</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$p < .05$.

Analysis of Attitudes toward mathematics learning

To uncover the attitudes toward learning mathematics (after the experiment), this study conducted multivariate analysis of covariance with respect to the perceptions of learners in three dimensions: learning enjoyment, learning motivation and anxiety caused by studying mathematics. The influence of the example integration method on attitudes toward mathematics was non-significant ($Wilks’ \, Lambda = 1.343$, $p = .241$, $\eta^2 = .047$, $Cohen \, d = .444$).

We also conducted analysis of covariance, as shown in Table 3. The influence of the example integration method on learning enjoyment was significant ($F_{(2,83)} = 3.815$, $p = .026$), indicating that learning enjoyment was greater among learners receiving TSCE ($M = 4.03$) than those receiving TCE ($M = 3.43$) or TSME ($M = 3.59$). No significant differences were observed in the learning enjoyment among learners receiving TCE or TSME. The influence of the example integration method on learning motivation and anxiety caused by studying mathematics were not significant (learning motivation: $F_{(2,83)} = .242$, $p = .786$; anxiety caused by studying mathematics: $F_{(2,83)} = .182$, $p = .834$).
Table 3. Summary of covariance analysis on attitudes toward mathematics learning

<table>
<thead>
<tr>
<th>Source of variance</th>
<th>Attitudes toward mathematics learning</th>
<th>SS (Type III sum of squares)</th>
<th>df (Degree of freedom)</th>
<th>MS (Sum of squares)</th>
<th>F (F test)</th>
<th>Sig. (Significant)</th>
<th>Post comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior knowledge</td>
<td>Learning enjoyment</td>
<td>.026</td>
<td>1</td>
<td>.026</td>
<td>.033</td>
<td>.855</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Learning motivation</td>
<td>.001</td>
<td>1</td>
<td>.001</td>
<td>.002</td>
<td>.969</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anxiety caused by studying mathematics</td>
<td>.572</td>
<td>1</td>
<td>.572</td>
<td>.882</td>
<td>.351</td>
<td></td>
</tr>
<tr>
<td>Example integration approach</td>
<td>Learning enjoyment</td>
<td>5.901</td>
<td>2</td>
<td>2.950</td>
<td>3.815*</td>
<td>.026</td>
<td>TSCE&gt; TSCE= TSME</td>
</tr>
<tr>
<td></td>
<td>Learning motivation</td>
<td>1.377</td>
<td>2</td>
<td>.688</td>
<td>1.293</td>
<td>.280</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anxiety caused by studying mathematics</td>
<td>1.184</td>
<td>2</td>
<td>.592</td>
<td>.913</td>
<td>.405</td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>Learning enjoyment</td>
<td>64.187</td>
<td>83</td>
<td>.773</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Learning motivation</td>
<td>44.207</td>
<td>83</td>
<td>.533</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anxiety caused by studying mathematics</td>
<td>53.855</td>
<td>83</td>
<td>.649</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

p < .05.

Discussion and conclusions

This study developed materials specific for teaching concepts related to equivalent fractions. Learning performance and attitudes toward mathematics were compared among students who had received TCE, TSCE, and TSME. Students in the TSME group benefited more than those in the TCE or TSCE groups with regard to basic inpainting, advance inpainting, and operational thinking. These results are in line with those presented by Lee and Chen (2009), indicating that the utilization of non-routine examples can promote learning performance in equivalent fractions. This may be explained by the fact that these non-traditional examples are an effective means to engage the intellect of students, capture their interest and curiosity, develop their mathematical understanding and reasoning processes, and allow for a variety of solution strategies, solutions, and representational forms. Unfortunately, elementary school textbooks in Taiwan present only traditional examples of equivalent fractions. To improve learning performance in equivalent fractions, alternative instruction models should be integrated into the curriculum and students should be encouraged to use non-routine examples.

The difference in learning performance between TCE and TSCE groups was not statistically significant. These results are consistent with those of Yuan, Lee, and Wang (2010), indicating that using virtual manipulatives can be as effective as using physical manipulatives. In other words, the form of the manipulatives makes little difference as long as the method of instruction is preserved, such that replacing the physical materials with virtual materials does not affect learning performance. In fact, our study found that the use of non-routine examples is the most important factor influencing the learning of equivalent fractions. Regardless of the instructional aid used, teachers should pay attention to the arrangement of worked examples in the application of virtual or physical manipulatives. Inappropriate examples can prevent students from learning how to solve problems associated with equivalent fractions.

Students in the TSCE group presented a more positive attitude toward learning mathematics, compared to those in the TCE and TSME groups. This echoes Reimer and Moyer’s (2005) study in which most students responded positively to the use of virtual manipulatives. However, students in TSME group did not describe greater learning enjoyment than those in TCE group, perhaps due to the difficulty and complexity of non-routine examples, compared with continuous examples. Thus, the design of worked examples should be considered an important factor affecting the attitudes of students when using virtual manipulatives.

In summary, this study found that learning outcomes with respect to equivalent fractions can be enhanced by the utilization of non-routine examples. We suggest that appropriate worked examples be provided during teaching, and virtual manipulatives be applied carefully in the instruction of mathematics in order to enhance learning enjoyment.
This study has a number of limitations. First, the sample size was small; therefore, these findings are not necessarily generalizable to other groups of learners with different educational or cultural backgrounds. Second, courses on equivalent fractions differ considerably from those of other domains such as biology or social sciences; therefore, our conclusions cannot be generalized to other disciplines. Finally, prior experience in areas such as mathematics epistemology or computer self-efficacy may influence learning outcomes when using virtual manipulatives or physical manipulatives. Future studies would need to examine the role of prior experience on learning with virtual manipulatives.

Acknowledgements

The funding of this study was supported by the National Science Council, Taiwan under grant NSC 101-2511-S-305-001-, NSC 102-2511-S-305-001-, NSC 103-2511-S-305-001-, NSC 100-2511-S-009-006-, and NSC 101-2511-S-009-006-MY2.

References


Learning Faults Detection by AIS Techniques in CSCL Environments

Amina Zedadra* and Yacine Lafifi

LabSTIC Laboratory, University 8 May 1945 Guelma, Algeria // BP 401, Guelma 24000, Algeria // zedadra_a@yahoo.fr // laf_yac@yahoo.fr

*Corresponding author

(Submitted August 22, 2013; Revised January 8, 2014; Accepted February 3, 2014)

ABSTRACT

By the increase of e-learning platforms, huge data sets are made from different kinds of the collected traces. These traces differ from one learner to another according to their characteristics (learning styles, preferences, performed actions, etc.). Learners’ traces are very heterogeneous and voluminous, so their treatments and exploitations are difficult, that make hard the tutors’ tasks. This paper introduces one of the bio-inspired computing techniques to improve the learning quality. In fact, Artificial Immune System (AIS) is a technique which was adapted for designing an assistant system that detects the wrong scenarios made by learners. Furthermore, this assistant system assists the learners in their activities. The main aim is to present the basic concepts of a new approach that aims at providing learners with relevant traces to improve their learning in order to minimize the tutor’s tasks. A novel algorithm is proposed to design the assistant system based on the two mechanisms of the AIS techniques (negative and clonal selection). The proposed algorithm was applied on a collaborative learning system called LETline 2.0 (http://www.labstic.com/letline/). An experiment was conducted in an Algerian University. The obtained results from this experiment were good and very efficient. The proposed approach enhances the cognitive and behavioral profiles of learners. In fact, the results show that the cognitive profiles of most students were improved. Also, it minimizes the tutor’s tasks.

Keywords

CSCL, AIS, Negative selection, Clonal selection, Assistant system

Introduction

In distance learning context, learning systems provide an environment where the learners can learn and interact with each other. These systems cultivate the abilities and the comportments of learners. However, this doesn’t mean that distance learning is perfect since it has many drawbacks, whereby a big number of learners are not successful at guiding their own learning process.

Every system has its own interface with heterogeneous content and different presentations of information. This fact leads to the emergence of a big problem in making the balance between the abilities of learners and the sight of the designers when developing the predictive scenarios. According to the heterogeneity of the interfaces, many learning systems cannot respond to the learners’ need and may waste time in wrong scenarios. Many studies (e.g., Djouad, 2011; Lafifi, Halimi, Herkas, Ghodbani, & Salhi, 2009; Sehaba, 2012) have revealed the use of learner’s traces (individual and collaborative activities among the users) to solve the problems mentioned above in order to boost the effectiveness of the collaborative learning systems. These traces are the results of the interaction between the actors, and the human actors and the system. However, traces are very voluminous and heterogeneous to be exploited manually by a human actor who is generally a tutor or a teacher. So, can the human actor manipulate all learners’ traces?

Our work context is the collaborative learning systems, where learners can learn and interact with each other to reformulate, test, refine, and repair their mental models (Hmelo-Silver, Jordan, Liu, & Chernobilsy, 2011). Also, it provides excellent opportunities for students to propose, support, evaluate, critique, and refine ideas in a more productive manner (Jeong, Clark, Sampson, & Menekse, 2011). In collaborative learning, learners need an additional support from human actors (teachers or tutors) to promote opportunities for effective collaboration. Besides, the collaboration in groups is difficult (Gweon, Jun, Lee, Finger, & Rosé, 2011), which confirms that an effective monitor and support is necessary to success the collaborative learning.

This research aims at providing answers to the following questions: Does the learner need an effective support and monitoring to improve his/her learning process? Is the tutor present usually when learners need monitoring? Is the
development of an assistant system for supporting and monitoring students can replace the human actor and can it success the collaborative learning?

We try to answer these questions in this research work. The latter belongs to the context of analysis and assistance of the learning situations to encourage students based on their traces. It aims at implementing an assistant system and integrating it in a CSCL (Computer-Supported Collaborative Learning) environment to improve the profiles of learners. For doing this task, the wrong scenarios must be found. For that, an algorithm of filtering traces is proposed to filter the irrelevant traces. Due to the problem context which aims at detecting the wrong scenario, bio-inspired techniques can be used. Indeed, the bio-inspired techniques are used in different fields: the anomaly detection in the computer security, detection of the application faults, optimization, pattern recognition, approximation functions, etc. More specifically, the Artificial Immune Systems (AIS) have important characteristics: detection, learning, memorization and cloning. So why not use the AIS techniques for the detection of the wrong scenarios. But, how can we adopt the AIS techniques to our problem? And is its integration gives good results?

The paper is organized in six sections. The section two presents the literature review. In section three, we give a brief overview of the AIS techniques in the field of e-learning and CSCL. Section four presents the architecture of the proposed system. In section five, we give the obtained results and the analysis of these results. Finally, section six presents a conclusion and the future works.

**Literature review**

Traces are defined in different ways according to the context of their application. Therefore, they have various definitions. Jermann and his team (Jermann, Soller, & Muehlenbrock, 2001) defined a trace as “an observation or a recording of the interaction of learners with a system for an analysis.” In 2005, another definition was proposed by Pernin (2005) where he defined the trace as “an index of actors’ activity, in an instrumented or not instrumented learning situation.” In a different way, Setoutti (2011) defined a trace as “an object collection, which is collected from an observation.”

According to the definitions given by previous authors, we propose the following definition that considers the trace as “a sequence that is defined by a series of actions done by the user when interacting with an environment.”

Traces are exploited to conceive: (1) analysis and assistance systems of learning situations which aimed at improving the task of actors monitoring (followed learners in their learning tasks and tutors in their different tasks, (e.g., Bousbia, 2011; Heraud, France, & Mille, 2004; Loghin, 2008), (2) engineering/reengineering of learning devices systems according to the information collected during the learning scenario in order to improve their quality (e.g., Diagne, 2009; Hussaan, Sehaba, & Mille, 2011; Luengo, Vadcard, Dubois, & Mufti-Alchawafa, 2006), and (3) adaptation/personalization of learning environment systems which adapt the contents of the learning platforms to the users (e.g., Guettat, Chorfi, & Jemni, 2010; Settoutti, 2011).

In the context of learning improvement based on learner’s traces, Heraud and his team (Heraud et al., 2004) proposed Pixed, a research project which uses the logs of learners’ interactions. These interactions are joined together as learning episodes to help the learners who were trying to find their learning paths. In another approach, France and her co-authors (France, Heraud, Marty, & Carron, 2007) introduce the traces’ visualization interface by elaborating the ClassroomVis system. The latter allows the tutor to observe and adapt learner group’ activities. Loghin (2008) regulates the learner’s activities by developing an observation station, which allows the collection of the student’s traces based on multi-agents systems. Diagne (2009) has revealed the re-use of the indicators that provide the tutor with information to control the learning activities on the cognitive, pedagogical, social and technical level. Compared with the previous approach, Bousbia (2011) has used also the indicators calculated from the learners’ navigation but to provide the teachers not the tutors with a perception of their learners’ behavior and to identify their learning styles for assisting them. In a different approach, the main aim of the work of Sehaba (2012) is to develop an adaptive helping system based on interaction traces. It consists in considering the traces left by the users as knowledge sources that the system can use to generate adapted help to the target user.

In the studies cited above, the process of assistance is made by the tutor and posteriori which is a major problem when the learners need a help and the tutor is not present online to help them. To solve these problems, we propose
an assistant system to improve the learning situation of learner in real time, by assisting the learners who follow a wrong scenario.

Other studies have the same aim of improving the user’s abilities in different fields, such as information retrieval (e.g., Fitchett, Cockburn, & Gutwin, 2013; Mele, 2013), Computer-supported cooperative work (e.g., Tausczik & Pennebaker, 2013), serious games (e.g., Bououd & Boughzala, 2013), pair programming (e.g., Radermacher, Walia, & Rummelt, 2012), etc. Table 1 gives a summary of the related works.

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
<th>Studies sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance learning</td>
<td>Helping learners in their learning processes</td>
<td>(Bousbia, 2011; Diagne, 2009; France et al., 2007; Heraud et al., 2004; Loghin, 2008; Sehaba, 2012)</td>
</tr>
<tr>
<td>Pair programming</td>
<td>Increase student performance in programming</td>
<td>(Radermacher et al., 2012)</td>
</tr>
<tr>
<td>CSCW</td>
<td>Improving the collaborative work</td>
<td>(Tausczik &amp; Pennebaker, 2013)</td>
</tr>
<tr>
<td>Serious games</td>
<td>Enhance the collaboration skills for a perfect team’s management</td>
<td>(Bououd &amp; Boughzala, 2013)</td>
</tr>
<tr>
<td>Information retrieval</td>
<td>Improve the research results and minimizing the time spent on retrieving</td>
<td>(Fitchett et al., 2013; Mele, 2013)</td>
</tr>
</tbody>
</table>

In recent years, many researchers have recognized the value of the use of AIS techniques in the field of Internet and web technologies. Artigues (Artigues, 2009) applies the AIS in the field of Virtual Reality Environments for Human Learning where he develops an intelligent tutoring system, which is based on the action recognition from users’ traces. In the field of web mining, Nasir and his colleagues (Nasir, A. Selamat, & H. Selamat, 2009) develop a system WMAIS (Web Mining using Artificial Immune System) that aims at recommending interesting and applicable news on the politics in Malaysia for the users. In another context, Romero and Nino (Romero & Nino, 2007) use the AIS for the extraction of the keywords from a document or a set of documents to enhance information retrieval. Also, AIS techniques are used in the field of Internet security (Chang, Venkatasubramanian, West, & Lee, 2013), spam filtering (Caruana & Li, 2012), pattern recognition (Hunt & Cooke, 1996) and anomaly detection (Gonzàles, 2003). According to the presence of these works, the current paper appears to be promising to develop and improve the models and algorithms using AIS techniques in e-learning.

Overview of the AIS techniques

Basic concepts

The field of Artificial Immune System has derived inspiration from many elements of the natural immune system to develop systems that operate in environments with constraints similar to those faced by the immune system (Hart & Davoudani, 2009). De Castro and Timmis (De Castro & Timmis, 2002) define the AIS as “the adaptive systems, inspired by the theories of the immunology, as well as the functions, the principles and the immune models, in order to be applied to the resolution of problems”.

The immunity is sub-divided into two distinct systems: innate immune system and adaptive immune system. The adaptive immune system has three principal processes (Timmis, Hone, Stibor, & Clark, 2008): negative selection, clonal selection and immune network. Whereas, Natural Dendritic Cells are the link between the innate and adaptive immune system.

Negative selection

The purpose of negative selection is to provide tolerance for self-cells (Aickelin & Dasgupta, 2005). The thymus is a gate against the non-self-antigens. The T cells presenting non-self-antigens are destroyed in this organ. All T cells retiring of the thymus and circulating in the body are said tolerantly towards the self.
Clonal selection

Clonal selection algorithm (De Castro & Von Zuben, 2000) is used by natural immune system to define the basic features of an immune response to an antigenic stimulus. It establishes the idea that only those cells that recognize the antigens are selected to proliferate. The selected cells are subject to an affinity maturation process, which improves their affinity to the selective antigens. The readers can read (De Castro & Von Zuben, 2002) for more details about the main stages of the clonal selection.

Immune network

Immune network theory is originally proposed by Jerne (Jerne, 1974). An artificial immune network is a bio-inspired informatics model that uses the ideas and the concepts of the immune network theory mainly the interaction between B cells and the cloning process. It receives an antigen as entry and sends back an immunized network compound of the B cells that are adjusting between them.

The immune network process is almost the same that the clonal selection, except that there exists a mechanism of deletion that destroys the cells having a certain threshold of affinity between them.

Dendritic Cells (DCs)

Dendritic Cells Algorithm (Greensmith, Aickelin, & Tedesco, 2010) is a second generation algorithm based on an abstract model of natural dendritic cells (DCs). It was introduced in 2005 (Greensmith, Aickelin, & Cayzer, 2005). Natural DCs are part of the innate immune system and are responsible for initial pathogen detection, acting as a vital link between the innate and the adaptive system.

AIS techniques in CSCL systems

In CSCL context, group interactions allow to improve the learners’ competencies through collaborative production, and break the insulation process by gathering learners using various communication tools: email, forum, blog, etc. CSCL systems provide users with collaborative tools that allow synchronous and asynchronous exchange between different actors. So, we have chosen to apply our algorithm in this field due to the great quantities of traces that are provided. The immune system uses many strategies to protect the body against all foreign molecules (non-self or antigen). According to its different features: mechanisms, self and non-self-discrimination, memorization, destruction of antigen, and so on, AIS techniques has been applied in different fields: computer security, optimization, pattern recognition, etc. Hence, we have chosen to apply the AIS techniques in the field of collaborative learning.

Our proposed solution is to develop a subsystem for detecting learning faults and assisting learners based on AIS techniques. The subsystem uses the negative and clonal selections to detect the irrelevant traces in order to improve the cognitive and behavioral profiles of learners.

Proposed filtering method

According to the difficulties and the limitations presented in the introduction and on the basis concepts of the AIS techniques (negative and clonal selection), we describe in this section our system architecture and our novel algorithm of filtering.

System overview

The principal thematic of our work concerns the support required to the human actors to improve learning quality. The fundamental idea is detecting the irrelevant traces and helping learners by identifying wrong scenarios using AIS
techniques, especially negative and clonal selection. The analogy between the natural immune system principle and the problem proposed above (Table 2) prompted us to develop our system.

<table>
<thead>
<tr>
<th>Natural Immune System</th>
<th>Artificial Immune System applied in our context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self (antibody)</td>
<td>Relevant trace</td>
</tr>
<tr>
<td>Non-self (antigen)</td>
<td>Irrelevant trace</td>
</tr>
<tr>
<td>Lymphocyte (B- and T-cell)</td>
<td>Detectors</td>
</tr>
<tr>
<td>Detect the antigen</td>
<td>Detect the irrelevant trace</td>
</tr>
<tr>
<td>Tolerization</td>
<td>Negative selection mechanism</td>
</tr>
<tr>
<td>Cloning and memorization</td>
<td>Cloning and update the self-data</td>
</tr>
</tbody>
</table>

**Overall system architecture**

As it is indicated by several authors, the treatment of traces passes by three main steps: collection, analysis and exploitation. Our context of work is situated in the second and the last step. Indeed, in the second step (i.e., traces analysis), the traces collected in the first step are filtered. In the last step, the aim is to recommend learners with the relevant traces and to provide the tutors with a small set of learner’s traces. Figure 1 illustrates the system architecture and its different functionalities.

---

**Collection of traces**

The system collects all learners’ actions. Traces can be resulting from the interaction learner/learner and learner/system. Also, other traces can be resulting from a collector of traces. We propose a novel classification of the types of traces, where we have divided them into five classes (Figure 2):

- Identification information: first name, family name, age, birthplace and class.
- Activities: learning activities, assessment and search for learning objects.
- Interaction and collaboration: assistance, synchronous collaborative tools, asynchronous collaborative tools and virtual meeting.
- Use: navigation on platform and on computer.
- Formation time: session’s duration, presence, habitual time for connection and consultation type (superficial, medium and deep).
Analysis of traces

- **Reformulation**: from the data set generated in the first step, a new data set is created where the traces will be reformulated according to the proposed format. The general format of the trace is defined as follows (Figure 3):

  \[ T = (A_1, A_2, \ldots, A_n) \]

  Each action \( A \) is characterized by the following 5-tuples:
  \[ A = (t, N, D, H_B, H_E) \]
  - \( t \): Type of the trace.
  - \( N \): Number of traces by week/month/year.
  - \( D \): Date of the trace.
  - \( H_B \): Trace beginning Hour.
  - \( H_E \): Trace end Hour.

- **Fusion**: in order to obtain all the traces left by learners during their use of the system, a fusion step is proposed. It consists in obtaining a hybrid traces having all the recorded traces according to their time (Figure 4).
• Cleaning: to eliminate the noise (false response from the server for example: page not found, error 404), an algorithm of cleaning traces is proposed (see Algorithm 1).

Algorithm 1: Cleaning traces

**Input:** $T$ = a set of traces
$NT$ = {error 404, erreur 404, page non trouvée, page not found, etc.} //It depends on the browser language (English, French, etc.).

**Output:** $CT$ = a set of cleaned traces

**Begin**

For all (trace $t \in T$) then
  For all (trace $nt \in NT$) then
    If ($t$ matches $nt$) then
      Discard $t$
    Else
      Place $t$ in $CT$
    End if
  End for
End

• Modem (modulator / demodulator): before the step of filtering, the modem is used as a modulator to convert the traces into scenarios (sequence of actions: $A \rightarrow B \rightarrow C$). Then, after the filtering step, the modem is used as a demodulator to transform the modeled trace into the original trace format according to the model proposed in the reformulation step (see Algorithm 2).

Algorithm 2: Modem process

**Input:** $CT$ = a set of traces (login to the system, learning, access to material, visualizing a resource, downloading resources, access to search engine, assessment, access to assessment, communication, send email, receive and answer email, access to the forum, a question on the forum, ask reply on the forum, request for assistance, receive a response to the request, response to requests for assistance, visualizing traces).

**Output:** $CTN$ = $CT$ in a novel format

**Begin**

For all (trace $ct \in CT$) then
  If ($ct$ = “Login to the system”) then
    $ct$ = “A”;
  Else if ($ct$ = “Learning”) then
    $ct$ = “B”;
    ...
  Else if ($ct$ = “Assessment”) then
    $ct$ = “G”;
    ...
  Else if ($ct$ = “Request for assistance”) then
    $ct$ = “O”;
    ...
End if
End

End
• **Filtering**: to detect automatically the wrong scenarios made by the learners during their learning sessions, we used an algorithm for filtering traces and keeping only the relevant ones. This algorithm is given in the next section (cf. Algorithm 3).

**Exploitation of traces**

The set of relevant traces is exploited by different actors:

- **Learner**: assistance and recommendation for learners when using the system (learning and collaboration).
- **Tutor**: minimize the tutors’ tasks by reducing the number of traces to visualize each time. We remember that the tutor has the possibility to visualize many types of traces left by learners every day, every week or during indicated dates. Furthermore, relevant traces are used to decrease the big number of the assistance requests sent by learners.

**The generic algorithm**

The pseudo code of the filtering algorithm is given in Algorithm 3 and its general structure is given in Figure 5. The algorithm has as input the self-data (a set of identified relevant traces) and gives us an output a set of irrelevant traces. In our context, the self-data are the relevant traces. The irrelevant traces are destroyed in order to let only the relevant ones that are used for recommending the learners to improve their learning scenarios. In fact, we use two mechanisms of AIS: negative and clonal selection. The negative selection is used to recognize the irrelevant sequences of actions. The clonal selection improves recognition and allows updating the initial database due to the memory cells. If a new trace arrives and it matches with the sequence of the initial, then the system made a clone, and it updates the initial base. Otherwise, it is an irrelevant trace; the system removes it and recommends learners with the relevant traces. For verifying if two traces match, we have used r-contiguous matching (Forrest, Perelson, Allen, & Cherukuri, 1994). In our case, r is equal to 3 because we have divided the sequence into equivalent sub-sequences that are composed of 5 actions.

**Matching with r-contiguous:**

We use the match between two strings which is proposed by Forrest and his colleagues (Forrest et al., 1994). The main aim of this matching is the following: when we have two strings x and y, match (x, y) is true if x and y agree (match) at least r contiguous locations.

**Example of matching:**

X: AMINAZEDADRA  
Y: AMINEABDAOU1  
Match (x,y) false in the case where r = 5 or greater.  
Match (x,y) true in the case where r = 4 or less.

---

**Algorithm 3: Filtering traces**

| Input: R=a set of identified relevant traces (R=CTN (output of algorithm 2)) |
| Output: IR=a set of detected irrelevant traces |
| Begin |
| Create an empty set D of detectors |
| Generate random traces $T_R$ |
| For all (trace $t_R \in T_R$) then |
| For all (relevant trace $r \in R$) then |
| If ($t_R$ matches $r$) then |
| Discard $t_R$ |
| Else |
| Place $t_R$ in D |
| End if |
Presentation of some interfaces of the realized system

In order to validate our idea, we have implemented a subsystem called AIS4FT (Artificial Immune System for Filtering Traces). It aims at filtering the set of traces to let only the relevant ones. Also, the system assists the learners by sending recommendations’ messages to improve their learning profiles. Figure 6 presents the number of daily detected wrong scenarios (before and after application of the filtering algorithm), while Figure 7 shows an example of a message that was sent automatically to a learner after detecting that he follows a wrong scenario.

From Figure 6, the system filters the irrelevant traces (big number) to a small number of only relevant ones. Also, the figure shows that the communication and the collaboration between learners are low in the first few days, whereas they are high in the middle of the semester and they begin to decrease in the last few days. As it is shown in the figure, between March 16th and March 31st, the interaction and the collaboration had reached more than 140, while there are only 25 relevant traces. During this period, the students were on holidays.

Figure 7 shows an example of recommendation message sent to a learner who following a wrong scenario. In fact, as it is shown in the figure, a learner had received a message having as subject: “recommendation”. It contains some
directions to follow in order to success his learning process. This recommendation message was sent automatically by the system after the application of our algorithm.

**Figure 6.** Number of bad scenarios before and after application of the filtering algorithm

**Figure 7.** Recommendation message sent to a learner

### Experiment: Results and discussion

#### Data set

In order to validate our approach we have used a set of traces from LETline system. LETline was developed by Lafifi and his team (Lafifi, Azzouz, Faci, & Herkas, 2010). It is an online learning and tutoring environment, which is composed of two important parts. The first one is a learning management system, which offers the teachers all the means for preparing their courses. Furthermore, the students are brought to build their knowledge. The second one is a tutoring system. It offers mainly all the assistance and capture tools, analysis and traces visualization to allow the follow-up of the students by a tutor.

In order to validate our filtering algorithm, we have added a new module to LETline for supporting collaboration. In fact, LETline 2.0 is the new version of LETline by supporting collaborative learning. As a result, LETline 2.0 is a CSCL system. So, we have integrated two tools into the system: the first one provides the traces resulting from learner/learner and learner/system interaction while the second tool gives a report about the interactions of the learner on his computer. The proposed algorithm is applied on traces collected from LETline 2.0 (see Table3).

<table>
<thead>
<tr>
<th>Table3. Dataset of collected traces</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary traces</td>
</tr>
<tr>
<td>5700</td>
</tr>
</tbody>
</table>
Participants

An experiment was conducted at Computer Science Department at Guelma University (Algeria). The participants were teachers, tutors and students of 2nd year (specialty: information systems). All these participants can use the platform at http://www.labstic.com/letline/ from any computer. The number of the actors participating in this experiment is shown in Table 4.

| Table 4. Number of human actors who used the LETline 2.0 system |
|------------------|------------------|------------------|
| Learners         | Tutors           | Teachers         |
| 60               | 10               | 2                |

Methodology

Students and tutors can access to the LETline 2.0 platform after a successful enrollment. Each tutor has his own groups (from one to five groups based on their preferences). The assigning of tutors is performed by the administrator manually.

The experiment was conducted in two stages: the first one was done without using the AIS4FT subsystem, while the second stage was done using the AIS4FT subsystem. During the experimental period, the students learn the concepts of “Languages Theory” subject. Two questionnaires were proposed by the teacher who is the responsible of the subject. These questionnaires contain a set of questions about two learning objects “languages” and “regular expressions”. Each questionnaire is composed of a set of Multiple Choice Questions (MCQ).

In the first stage of the experiment, a pretest (first questionnaire) was given to evaluate the cognitive profiles of the learners without an assistant system. However in the second stage, a post-test (second questionnaire) was given to measure the usefulness of the proposed subsystem on the cognitive profiles of the learners. The cognitive profiles were calculated basing on the results of the questionnaires (responses of the students on the proposed questions). We adopted the same formulas proposed by Lafifi and his team (Lafifi et al., 2010).

We have chosen ten students randomly to verify the usefulness of the recommendation proposed to the students when interacting with each other, and for verifying our experiment hypothesis. During the system use, the students can visualize their cognitive and behavior profiles. Also, the tutors can visualize the profiles of their learners to assist them.

Results and discussion

The hypothesis of the research is:

- Null hypothesis H0: the use of the AIS techniques cannot help students to improve their cognitive profiles (learning process).

In order to verify our hypothesis, we calculated the difference of the cognitive profiles of the students. For doing this, each student must answer some assessment questions.

As we have the same sample of students who used the system in two phases (before and after) and the number of the learners is less than 30, we used student paired t-test. Before using the paired t-test, we must ensure that the data are normally distributed. For doing this, we have used Shapiro-Wilk test. The results obtained in both samples are: pretest data: $W = 0.86$, $P = 0.07$; post-test data: $W = 0.95$, $P = 0.74$. Since $P_{value}$ in both samples is greater than the significance level of $P = 0.05$, our data sets were normally distributed.

To know if the difference is significant between the two means, a paired t-test is appropriate to compare how a group scores vary in two different test conditions (sample size is equal to ten). R software is used to calculate the statistical data. The following results are obtained with a confidence level of 95% ($\alpha = 0.05$).
From the Table of t-test, $t_{0.975} = \pm 2.26$, so $t_{score} > t_{0.975}$ ($4.94 > 2.26$). As results, the difference was statically significant; therefore the null hypothesis H0 is rejected. So, the alternative hypothesis is proved and we can affirm that “the use of AIS techniques improves the cognitive profiles of learners”.

The main aim of this research was to investigate the effect of our AIS4FT subsystem on the students’ profiles. Firstly, there was a significant difference between the pretest mean of the students ($Mean = 2.47; SD = 1.80$) and the post-test mean of the students ($Mean = 5.81; SD = 2.23$). Secondly, Pearson’s correlation test was also performed to analyze the correlation between the students’ scores. In terms of the correlation coefficient, the size of the observed effect ($r = -0.63$) indicates that the students’ scores are negatively correlated. From the obtained results, we observe that our system detects automatically the wrong scenarios in real time and improves them.

In brief, we can say that the use of AIS techniques to detect learning faults can enhance the cognitive profiles of learners. Concerning the collaboration and the dynamism between the learners, we have calculated their behavioral profiles. We found that the behavioral profiles of most students were improved. These profiles are calculated from the number of asynchronous tools used in the communication and collaboration processes.

**The effects of the proposed algorithm on detecting the wrong scenarios**

Figure 8 presents the variation of the scenarios of ten learners when they are connected to the LETline 2.0 platform for the first days.
As it is shown in Figure 8, the abscissas are the different traces and the ordinate are the order of traces in the scenario of each learner. Each trace has its position in the scenario from one to eighteen (e.g., all the learners login to the system which is the first trace in the scenario, after that each learner does his own activities). We observe that the variation of learner’s scenarios is not perfect, that are resulting from the wrong navigation of learners that let them waste time with wrong scenarios (For example, learner 8 logsins to the system first (1st position), access to the learning space at the end of his scenario (18th position), access to the forum (14th position), assessment (6th position), which gave a bad results with low profiles).

Figure 9 presents the variation of learner’s scenarios after applying our algorithm.

As it is presented in Figure 9, we observe that the learner’s scenarios are improved in a good way, where the subsystem of filtering detected the wrong traces and recommend learners with the right scenarios. In Figure 9, we observe that eight learners (learner 1, learner 2, learner 3, learner 4, learner 6, learner 7, learner 9, and learner 10) improve their scenarios using the recommendation provided by the subsystem, which includes the relevant traces to do (actions). In fact, the scenario of the learner 5 and the learner 8 is not improved because they didn’t take into consideration the recommendations proposed by the filtering subsystem.
Finally, we can confirm that the introduction of AIS techniques in CSCL systems for detecting learning faults gives good results.

**Conclusion and future work**

Immune system theory is a natural defense against the foreign molecules. It has different features such as detection of anomaly, learning, memorization, cloning and adaptation. For these reasons, the immune system can be used as a mechanism for inspiration in a variety of fields. According to the different characteristics of the AIS techniques, there are several implications for including AIS theory in CSCL systems. This inclusion aims at improving the learning quality and supporting students by providing several advices for them.

As one of the AIS techniques feature is the detection of anomaly, we have adapted this characteristic to our problematic, which is the filtering and the improvement of the wrong scenarios. A wrong scenario (set of irrelevant traces) is the sequence of actions that conducts to low levels of learners.

In this paper, we have adapted and applied AIS techniques in a CSCL system. The main aim of our work is to present to the learners their irrelevant traces (which returns to the problems of the bad navigation of learners where they are wasting a long time to understand the platform functionalities) and provide them with those relevant. The goal is to improve learners’ scenarios in real time and provide the tutors with a minimized set of traces to help them in their different tasks. The filtering process uses AIS techniques where the negative and clonal selections are combined. The negative selection is used to detect the irrelevant traces and the clonal selection is used to update the initial base. The developed subsystem detects all the irrelevant sequences of actions, and recommends learners to improve their learning process.

In order to verify and examine the proposed approach, some typical applied tests based on the LETline 2.0 platform were introduced and explained. From the experimentation performed by students from an Algerian University, we note that the majority of learners solves their problems without the intervention of the tutors. Tutors could be absent due to some unexpected reasons. So, it is legitimate to the tutors to don’t be online all the times when an assistant system is used to recommend the learners with the relevant traces in order to complete their learning sessions successfully.

With the increase of traces that are recorded on the database, the initial database is also increased. So, the execution of the algorithm will take a long time for filtering the traces, thus our future work is to optimize the complexity of the proposed algorithm. Whereas, the problem isn’t in the algorithm, but in the size of the initial base that is updated every time. To this end, we propose to classify the sub-sequences into several classes where the test will be made only on a subclass of the initial base and not on the whole database. The second future work is the interpretation of traces in CSCL systems. Furthermore, we propose to implement a search engine for finding the relevant traces.

**Acknowledgments**

The authors thank all those who have participated in the experiments in real situations (teachers, tutors, and students of Guelma University). Also, the authors thank the reviewers for their important suggestions for improving the quality of the paper.

**References**


Role of Parents and Annotation Sharing in Children’s Learning Behavior and Achievement Using E-Readers

Wu-Yuin Hwang¹, Yi-Fan Liu¹, Hon-Ren Chen²*, Jian-Wun Huang¹ and Jin-Yi Li³

¹Graduate School of Network Learning Technology, National Central University, Taoyuan 320, Taiwan // ²Department of Digital Content and Technology, National Taichung University of Education, Taichung 403, Taiwan // ³Chyau-Zhong Elementary School, Taichung 427, Taiwan // wyhwang@cc.ncu.edu.tw // yifan.liu.tw@gmail.com // hjw85296@gmail.com // hrchen@mail.ntcu.edu.tw // miniljy@gmail.com

*Corresponding author

(Submitted August 6, 2013; Revised January 2, 2014; Accepted March 11, 2014)

ABSTRACT

Although previous studies have highlighted the advantages of using e-books for learning, most have compared learning achieved with traditional textbooks with that achieved with e-books in a classroom situation. These studies focused on individual learning instead of on interactions among learners, learning behavior using e-books after school, and parental observations of children using e-books for learning. This research investigated the use of annotatable multimedia e-readers (AMEs) for elementary school-level English and examined the effects of annotation-sharing mechanisms inside and outside the classroom on learning and achievement. The research findings suggest that reading the annotations of high-achieving learners (HLA) via the annotation-sharing function can reinforce learning. Moreover, HLA annotate significantly more in class and after school than do low-achieving learners (LLA). We found a positive correlation among parents’ perceptions of after-school learning with AMEs, learners’ after-school annotation recording, learning behavior while listening to such recordings, and children’s learning achievement.

Keywords

Annotatable multimedia e-readers (AMEs), Annotation sharing, English learning, Learning behavior, Learning time

Introduction

Electronic readers and smart phones have fast become a part of our daily lives. In particular, the convenience of these devices and the potential for enhancing the user’s performance through their use is extraordinary. In terms of learning, e-readers allow learners to access material anywhere and anytime. In addition, e-readers with rich multimedia tools and annotation capabilities, such as those offered by Google, also assist in the learning process. Previous research has indicated that e-readers could be used to improve the quality of learning (Huang, Liang, Su, & Chen, 2012; Korat, 2010; Korat & Shamir, 2012; Li, Chen, & Yang, 2013; Shih, Chen, Cheng, Chen, & Chen, 2013). Li et al. (2013) developed an e-reader system with a visual cue map to help learners construct cognitive maps effectively. Huang et al. (2012) developed an interactive learning system with e-readers for elementary school students; their principles for designed multimedia interactive learning were adopted to help students’ personalized learning experiences with e-readers. These studies concluded that the proposed interactive learning system could create a better personalized learning experience for elementary school students.

Recently, annotation has become a popular activity in the learning process (Chen, Hwang, & Wang, 2012; Hwang, Shadiev, & Huang, 2011; Hwang, Wang, & Sharples, 2007) because annotation is associated with the output of knowledge internalization (Peverly et al., 2013). To effectively improve learners’ performance, annotation tools or activities should be added to e-readers or smart phones. Previous research has also indicated that learners using annotation can effectively improve their performance (Hwang, Chen, Shadiev, & Li, 2011; Hwang & Hsu, 2011). However, annotation belongs to personal learning and usually has personal meaning. The annotations of different learners are usually not the same. It is important to enhance peer interaction based on annotation sharing so that learners can easily review others’ annotations and exchange ideas. Research has indicated that the annotation-sharing mechanism can enhance peer interaction and learning motivations. Yang, Zhang, Su, and Tsai (2011) developed a semantic web-service-supported multimedia tool to facilitate collaborative annotation sharing in the context of computer-supported collaborative learning (CSCL). Their experimental results showed that annotation sharing can facilitate knowledge sharing and improve participating students’ reading comprehension. Su, Yang, Hwang, and Zhang (2010) also indicated that the annotation sharing mechanism can increase learners’ performance in...
collaborative learning environments. Moreover, their results showed that the influence of annotation on learning performance becomes stronger with the use of the sharing mechanism.

In addition to learning annotation, learning time (the amount of time students are actively engaged in learning) also has an effect on learning performance (Adelman, 1996; Black, 2002; Gettinger & Walter, 2012). Therefore, learning time is worth exploring. However, evaluating the effect of learning time only at school may not be enough. Students’ learning time and learning practices after school should also be considered. A key element that may enhance learning after school is parental involvement (Fan & Chen, 2001). Desforges and Abouchaar (2003) revealed that parental involvement not only improved children’s willingness to learn, but also improved children’s attitudes towards after-school learning. Hanna et al. (2011) showed that parental involvement can effectively enhance children’s learning performance. Thus, to facilitate learning performance, parental involvement with the student is essential, from an educational viewpoint (Sartor & Youniss, 2002).

Previous studies have mainly compared e-readers and traditional paper-based books (Huang et al., 2012; Korat, 2010; Li et al., 2013; Shamir, Korat, & Fellah, 2012). Wright, Fugett, and Caputa (2013) evaluated the use of reading resources available between the two reading methods, and found that vocabulary acquisition and reading comprehension were similar. Huang et al. (2012) indicated that using an e-reader or printed book made no significant difference in reading accuracy. Korat and Shamir (2012) examined the effect of e-reader use for children in the context of vocabulary and reading and found that those who read e-readers exhibited significant progress in reading and understanding the meaning of words, compared to the control group. However, all of these previous studies mostly compared the use of e-readers and traditional books in class, without investigating the effects of using e-readers after school. Parental involvement must be considered when children are allowed to bring e-readers home with them; however, few studies have addressed this issue.

To address this gap in knowledge, we investigated learning behavior when students used e-readers both in class and after class, the effect of e-readers on learning performance, and how parental involvement influences after-school learning. In addition, we compared the differences in learning behavior and learning time between in-class and after-class learning, and with and without annotation sharing between learners with a high level of learning achievement (HLA) and a low level of learning achievement (LLA). This research addresses several key questions related to learning behavior, learning practices, and the performance of students using e-readers: (a) what are learners’ perceived attitudes toward the AME system; (b) what are parents’ perceptions toward their children’s use of the AME system after class and the relationship between parental perception and their children’s learning behavior after class with respect to learning achievement; (c) what are the effects of the learner’s behavior in class/after class on their learning achievement; (d) is there any significant difference in learning behavior and learning time between HLA and LLA students, with/without annotation sharing in class/after class; and (e) what are the effects of reviewing peers’ annotations on learning achievement?

**Literature review**

**E-readers**

E-readers have renewed the traditional concept of the book, encouraging the socialization of reading and user participation (Cordon-Garcia & Lopes, 2012). Most recently, light and mobile reading devices with high-resolution displays have become popular and have opened new opportunities for reading applications in education, business, and the private sector (Siegenthaler, Bochud, Bergamin, & Wurtz, 2012). For example, a single electronic device can be used for a variety of activities, in addition to reading a book (Siegenthaler, Wurtz, Bergamin, & Groner, 2011). More specifically, e-readers offer several major advantages, such as portability, searchability, and multimedia characteristics (Shurtz & von Isenburg, 2011). These characteristics have made e-readers very popular and useful in educational settings.

Thus, the e-reader has received considerable attention from scholars in educational settings. For example, Korat (2010) revealed how e-books facilitated learning by offering automatic dynamic visuals that dramatized the story details for a more complete story scene. In this case, the researchers focused on developing a multimedia tool for educational e-books, which was aimed at supporting elementary students who had already begun the process of learning English as a foreign language (EFL) at school. Students in the experimental EFL group who used the e-book
showed a significantly improved learning performance compared to those in the control group (without e-books and the developed multimedia support). Huang et al. (2012) designed personalized learning tools, such as e-annotation, bookmarks, content searching, and learning-process tracking for e-books. Here, the individual’s learning needs and characteristics were emphasized. The usability and functionality of the proposed system were well-suited for most students, and students obtained a better personalized learning experience. Li et al. (2013) proposed a mechanism using visual contextual cues to facilitate students’ building of cognitive maps using e-books; these e-books offered contextual cues and navigational mechanisms. Participants who used the e-book system with the visual cue mechanism spent significantly less time completing ten navigational tasks and gained a higher reviewing score than those who did not use the visual cue mechanism.

Annotation

In e-learning environments, the effect of digital annotation may be as good as paper-based annotation with regard to learning performance. In one study, no significant difference in learning achievement was observed between note-taking on paper and note-taking in a learning system (Quade, 1996). Chen and Liu (2012) investigated the effectiveness of using learner-generated as opposed to instructor-provided multimedia annotations on foreign language reading comprehension and attitudes. Students in the learner-generated annotation group showed better reading comprehension than those in the instructor-provided group, which suggests that learner-generated annotation in e-learning environments could also support reading comprehension. Hwang, Liu, Liu, and Huang (2011) constructed the VPen system, which enables students to reflect on learning material individually by creating annotations and then collaboratively by sharing annotations with their peers or teacher for in-depth discussions of their thoughts and ideas. The system facilitated writing and speaking performance, and therefore improved learning achievement. Thus, the annotation system provided the support needed for students to improve their learning performance. Hwang et al. (2011a) noted that students’ actual learning behavior was correlated with their speaking and writing performance, which in turn was correlated with their learning achievement. Hwang et al. (2007) proposed personalized tools for individual learning systems and analyzed students’ learning behavior. The influence of annotation on learning performance became stronger with the use of sharing mechanisms.

Annotation-sharing mechanism

Knowledge sharing in CSCL requires intensive social interactions among participants, typically in the form of annotations (Yang et al., 2011). Students can review the annotations of their peers using an annotation-sharing mechanism. Previous research has shown the effects of the annotation on students’ ability to learn (Lee, Lee, & Leu, 2009; Su et al., 2010; Yang et al., 2011). Specifically, Su et al. (2010) demonstrated the use of a personalized annotation management system for managing, sharing, and reusing individual and collaborative annotations. Use of the system increased learning achievements in collaborative learning environments. Chen et al. (2012) found that interactivity and helpfulness were statistically significant factors for predicting the future use of an annotated system. In addition, note-taking habits also affected learners’ perceptions of using the annotated system. In summary, a significant body of research shows that the annotation-sharing mechanism supports the learner and helps improve learning performance.

Learning behavior and learning time

Students have to strike a proper balance between time spent learning and free time. Time spent learning should be used effectively to increase the level and amount of learning (Wyne & Stuck, 1982). Gettinger (1985) showed a decrease in learning and retention by 11–18.5% when insufficient time was spent on or allocated for learning. Fisher (2009) investigated the distribution of learning time for senior high school students, and found that the majority of time in school was spent listening or waiting; students were engaged in reading, writing, and peer work for only a small fraction of the school day. That study suggested that if learners wanted to improve their learning performance, they had to change the way they managed their time spent learning. Black (2002) advocated that students should spend their learning time more efficiently. In summary, two key factors affecting students’ learning performance are adequate learning time and the effective use of the time spent learning.
Importance of after-school learning

Language studies, in particular, require time spent learning after school to achieve better fluency, and studies have shown that after-school learning enhances learning performance in this area (Bergin, 1996). Noam, Biancarosa, and Dechausay (2003) divided after-school education into three major categories: (1) bridging school to after school, (2) homework or extended learning, and (3) additional curricula or enriched learning. Thus, the main purpose of after-school learning is to continue or extend learning of the content presented in the classroom. A suitable learning plan facilitates effective after-school learning.

There have been several studies on the effect of after-school learning on learners’ achievement (Dairianathan & Subramaniam, 2011; Noam et al., 2003; Tran, 2011). Tran (2011) examined the relationship between students’ out-of-school experiences and various factors associated with science learning. The ability to make connections between in-school and out-of-school science experiences was associated with positive learning outcomes, such as achievement, interest in science, and effort in learning science. Dairianathan and Subramaniam (2011) investigated primary students’ learning through participation in out-of-school activities. Participation in out-of-school activities not only improved learning performance, but also enhanced the motivation to learn.

Parental involvement

In the current educational setting, after-school learning has become increasingly important as developmental support for the learner. However, parental involvement may be a key factor affecting after-school learning (Krumm, 1996). There have been several studies on the effect of parental involvement on student learning outside of class (Desforges & Abouchaar, 2003; Floyd & Vernon-Dotson, 2009; Gonzalez-DeHass, Willems, & Holbein, 2005; Sartor & Youniss, 2002). Xu (2004) showed that learning attitude was greatly affected by family involvement with homework. Parental involvement not only improved children’s willingness to learn, but also improved their attitudes towards after-school learning (Gonzalez-DeHass et al., 2005). In summary, there is a body of evidence showing that parental involvement is key for effective after-school learning.

Methods

Participants and learning material

The participant consisted of 31 sixth graders (11–12 years old; 16 males and 15 females). All participants were provided with an e-reader for learning the English language. The learning materials contained in the e-reader were the same as those in the original textbooks.

Research design

This study adopted a one-group pretest–posttest quasi-experimental research method to investigate the effects of annotatable multimedia e-reader (AME)-based English instruction on elementary school students’ learning achievement and to explore differences in the in- and after-school learning behavior and learning time of students with different learning achievements. Questionnaires soliciting parents’ perceptions and learners’ feedback were based on the technology acceptance model (TAM) proposed by David (1986). TAM introduces two perspectives: perceived usefulness and perceived ease of use, which explain and predict users’ attitudes and the factors that contribute to the use of the new technology. TAM has been applied by empirical research predicting intentions to use new information technology tools (Venkatesh & Davis, 2000).

Learners used the AME system over five weeks. During the first two weeks, they used the system without annotation sharing. From the third through to the fifth week, they used it along with annotation sharing. During the formal experiment, e-readers were used as the main learning tool (i.e., students used the e-readers in class and were allowed to take the e-readers home after school). The experimental procedure is illustrated in Figure 1.
Research tool

Annotatable multimedia e-readers (AME)

The AME system was developed to support learning, providing learners with multiple learning resources, including text anchors, image anchors, voice annotation, the ability to hear their own annotation comments, a text-to-speech function, and the ability to hear the teachers’ annotations. This learning tool included the following features:

- **Text anchors.** Learners could tag, create line drawings, and annotate in a textual content, as shown in Figure 2. They also could modify the content of their text annotations at any time.

- **Image anchors.** Learners could annotate non-text content, as shown in Figure 3. In general, the main purpose of this feature was to support learners in adding supplementary words or sentences from their teachers’ lectures.

- **Voice annotation.** We developed a voice-recording tool for the AME system that enabled learners to record their pronunciation, as shown in Figure 4.

- **Text-to-speech (TTS).** Learners could use the TTS tool to support their understanding of how to pronounce words or sentences, as shown in Figure 5.

- **Dictionary.** We developed a multimedia dictionary tool to translate words or sentences quickly and to provide suitable sentences for learners. The multimedia dictionary tool could also perform word translations offline to support learning after class.

- **Teacher lecture annotation.** We developed a lecture-annotation feature to record the lectures given by teachers in class, enabling students to listen to them after class, as shown in Figure 6.

- **Parental signature.** The parental signature could be used as a communication link between parents and teachers. Thus, parents and teachers could have more opportunities to understand students’ learning progress or problems and could better monitor their learning in Figure 7.
Annotation sharing. The annotation-sharing mechanism can facilitate knowledge sharing and improve student comprehension (Yang, Zhang, Su, & Tsai, 2011) as well as increase learning achievement (Su, Yang, Hwang, & Zhang, 2010; Yang, Tseng, Shih, & Liang, 2012). Annotation sharing mechanisms can be divided into two types: full annotation sharing and group annotation sharing (Hwang, Wang, & Sharples, 2007). The former allows all students to review one another’s annotations, whereas the latter allows only group members to review one another’s annotations. We adopted the full annotation-sharing mechanism to obtain a more complete understanding of learning behavior, and we generated a list of students ranked by the number of annotations they made. Thus, students were able to review one another’s annotations (e.g., text and voice annotations) and improve their own annotations when they encountered problems or difficulties after class. More importantly, our annotation-sharing mechanism allowed students to review but not modify the annotations made by others. This may encourage students to make more annotations, as shown in Figure 8.
Parents’ perception questionnaire

We used paper-and-pencil questionnaires to investigate parents’ observations of their children’s after-school learning using the AME system. The questionnaire was verified by three experts in the e-learning field who revised or eliminated items with ambiguous meanings. Twelve questions remained, and these had a Cronbach’s $\alpha$ of 0.946. The questionnaire was divided into three sections: parental perceived ease of use, parental perceived usefulness, and learning behavior observed by parents. All questions were scored using a five-point Likert scale: strongly agree, agree, neutral, disagree, and strongly disagree. Definitions of the three sections are presented in Table 1. All questions are provided in the Appendix A.

<table>
<thead>
<tr>
<th>Parts</th>
<th>Operational definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parental perceived ease of use</td>
<td>Parents think that children can learn English more easily by using an AME system as the main learning tool.</td>
</tr>
<tr>
<td>Parental perceived usefulness</td>
<td>Parents think that the functions of the AME system help children learn English at home.</td>
</tr>
<tr>
<td>Learning behavior observed by parents</td>
<td>Parents identify children’s learning behaviors at home in terms of their after-school learning activities.</td>
</tr>
</tbody>
</table>

Learners’ feedback questionnaire

We solicited feedback from learners using a questionnaire investigating their thoughts about learning via an AME system. The questionnaire was validated by three experts in the e-learning field who revised or eliminated questions with ambiguous meanings. The final instrument consisted of 23 questions with a Cronbach’s $\alpha$ of 0.854. The questionnaire was divided into five sections: perceived ease of use, perceived usefulness, intention to use, perceived usefulness after school, and intention to use after school. All questions were scored using a five-point Likert scale: strongly agree, agree, neutral, disagree, and strongly disagree. Definitions of the five sections are presented in Table 2. All questions are provided in the Appendix B.

<table>
<thead>
<tr>
<th>Parts</th>
<th>Operational definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived ease of use</td>
<td>Learners think that the AME system is easy to use.</td>
</tr>
<tr>
<td>Perceived usefulness</td>
<td>Learners think that the AME system can enhance their learning in English.</td>
</tr>
<tr>
<td>Intention to use</td>
<td>Learners intend to use the AME system in class.</td>
</tr>
<tr>
<td>Usefulness after school</td>
<td>Learners think that the AME system can enhance their English learning after school.</td>
</tr>
<tr>
<td>Intention to use after school</td>
<td>Learners intend to use the AME system after class.</td>
</tr>
</tbody>
</table>

Data analysis

The data sources of this study were participants’ learning behavior and learning time by using AME system in class and after school in e-book learning and English learning scores of pretest and posttest. Data analysis was based on SPSS statistical software. Reliability analysis of questionnaire designed was validated by Cronbach $\alpha$ to test the internal consistency of the questionnaire. Pearson correlation coefficient was used to test the correlation among the dimensions and correlation between learners’ learning behavior and learning achievement. Simple linear regression was conducted to probe into the prediction of one independent variable on one dependent variable. The value of explanatory power can be found to realize if annotation behavior can predict learning achievement. We conducted a $t$-test of independent to find out if, with and without annotation sharing, learners of different learning achievements have significantly different learning behavior and learning time in class and after school. The interview script includes two primary opinions. First, parents want to be able to let their children use e-readers to learn English again. Secondly, learners agreed that e-readers, which provide a lot of multimedia tools, can help learners easily learn English.
Result

Questionnaire analysis

Result of parents’ perception questionnaire analysis

The data obtained from the questionnaires was applied to identify parents’ perceptions and attitudes towards their children use of the AME system. Table 3 shows the results of the mean score and standard deviation of each part of the parents’ questionnaire. The result shows that the most of parents agreed that their children can learning English by using the AME system.

The result of the parents’ open-questions also showed that the most of parents want to be able to let their children use e-readers to learn again. Most of parents made comments similar to the following:
- “My children use e-readers to learn again, because e-readers can learn easily when my children reviewing after class.”
- “I’m willing to let my children use e-readers to learn again, because e-readers can provide a lot of learning resources like teacher’s lecture annotation and text-to-speech.”

As a result, it was found that the most of parents prefer to let their children use e-readers for learning at home.

<table>
<thead>
<tr>
<th>Parts</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parental perceived ease of use</td>
<td>3.95</td>
<td>0.57</td>
</tr>
<tr>
<td>Parental perceived usefulness</td>
<td>4.10</td>
<td>0.52</td>
</tr>
<tr>
<td>Learning behaviors observed by parents</td>
<td>3.68</td>
<td>0.94</td>
</tr>
</tbody>
</table>

Result of learners’ feedback questionnaire analysis

Table 4 shows the means and the standard deviations of the learners’ perception. All items in parental perceived ease of use \( (M = 4.10) \), perceived usefulness \( (M = 3.95) \), perceived intention to use \( (M = 3.90) \), perceived usefulness after class \( (M = 4.04) \) and perceived intention to use after class \( (M = 3.81) \) of the AME system dimensions were ranked with a high score. It indicated that most of learners agreed that using the AME system was easy and useful.

The result of the learners’ interviews also showed that the most of learners agreed that e-readers, which provide a lot of multimedia tools, can help them easily learn English. Most learners made comments similar to the following:
- “When I want to review English at home, it is convenient to listen to the teacher’s lecture by using e-readers.”
- “E-readers are useful, because they provide a lot of functions, like text-to-speech, dictionary.”
- “I can take notes easily using e-readers.”

As a result, it was found that the most of learners agreed that e-readers can help them easily to learn English. This may be the reason why learners prefer to use e-readers to learn.

Table 4. The mean score and standard deviation of each part of the learners’ questionnaire

<table>
<thead>
<tr>
<th>Parts</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived ease of use</td>
<td>4.24</td>
<td>0.779</td>
</tr>
<tr>
<td>Perceived usefulness</td>
<td>3.95</td>
<td>0.910</td>
</tr>
<tr>
<td>Intention to use</td>
<td>3.90</td>
<td>0.778</td>
</tr>
<tr>
<td>Usefulness after class</td>
<td>4.04</td>
<td>0.959</td>
</tr>
<tr>
<td>Intention to use after class</td>
<td>3.81</td>
<td>0.845</td>
</tr>
</tbody>
</table>

Note. Correlation between parents’ perceptions and learning achievements.

As shown in Table 5, the results showed that learning behaviors observed by parents have significant correlation with quantity of voice annotation \( (r = .393^*, p < .05) \) and listening one’s own voice annotation \( (r = 0.458^{**}, p < 0.01) \). This implies that the learners made more voice annotations. Listening to own voice annotation, their learning
behaviors would more easily observed by their parents after the class. On the other hand, the results showed that the learning behaviors observed by parents have significant correlation with learning achievement \( (r = 0.424^*, p < 0.05) \). This implies that the learners’ learning behaviors more observed by their parents will have high learning achievement. In other words, the parents play an important role in their children’s learning English after class.

### Table 5. Correlation analysis of parents’ perceptions and learners’ learning behavior after class

<table>
<thead>
<tr>
<th>Learning behaviors observed by parents</th>
<th>TA</th>
<th>VA</th>
<th>OVA</th>
<th>TTS</th>
<th>TVA</th>
<th>LA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.255</td>
<td>.393*</td>
<td>.458**</td>
<td>.192</td>
<td>.218</td>
<td>.424*</td>
</tr>
</tbody>
</table>

*Note. TA = Text annotation after class; VA = Voice annotation after class; OVA = Listening of own voice annotation after class; TTS = Text-to-speech after class; TVA = Listening to a teacher’s voice annotation after class; LA = Learning achievement. * \( p < .05 \). ** \( p < .01 \).*

**Correlation between learning behavior and learning achievements in and after class**

This section presents the results that the correlation between learners’ learning behavior in and after class and learning achievement. Pearson’s correlations were used to examine the relationships between learning behavior and learning achievement. As showed in Table 6, the results shows that learning behavior in which learners made text annotations in class is significantly correlated with their learning achievement \( (r = 0.433^*, p < 0.05) \). It indicated that the more learners used text annotation in class, the better their learning achievement.

In order to examine whether the quantity of text annotation in class are significant predictors for learners’ learning achievement, we conducted a simple regression analysis. As showed in Table 7, the results demonstrate that quantity of text annotation in class were the significant predictors for students’ learning achievement \( (p < 0.05) \). It indicates that learning behavior in which learners made text annotations in class can facilitate their learning achievement.

On the other hand, the result of Pearson’s correlation analysis showed that learners’ learning behavior after class had no significant correlation with learning achievement \( (p > 0.05) \). It implies that the quantity of annotation after class may not be enough to predict learners’ learning achievement. As mentioned before, “learning behaviors observed by parents” have significant correlation with learning achievement. It implies that only the outputs of learners’ learning behavior after class could facilitate their learning achievement (Fisher, 2009).

### Table 6. Correlation analysis of learning behavior in/after class and learning achievement

<table>
<thead>
<tr>
<th>Learning achievement</th>
<th>In class</th>
<th>After class</th>
<th>In class</th>
<th>After class</th>
</tr>
</thead>
<tbody>
<tr>
<td>TA</td>
<td>.433*</td>
<td>.149</td>
<td>.257</td>
<td>.322</td>
</tr>
</tbody>
</table>

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

### Table 7. Simple regression analysis for learning achievement

<table>
<thead>
<tr>
<th>Model</th>
<th>Predicting variables</th>
<th>B</th>
<th>SE</th>
<th>Beta</th>
<th>t</th>
<th>R2</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning achievement</td>
<td>Text annotation in class</td>
<td>.223</td>
<td>.086</td>
<td>.433</td>
<td>2.583</td>
<td>.159</td>
<td>.015</td>
</tr>
</tbody>
</table>

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

**The effect of sharing to learning behavior and learning time in and after class**

We classified the participants into two groups based on their average score gained from their post-test. More specifically, those who gained higher scores than the average were assigned to the HLA group \( (N = 15, M = 95.13, SD = 3.16) \), whereas those who gained lower scores than the average were assigned to the LLA group \( (N = 16, M = 64.56, SD = 19.08) \). Furthermore, this section focuses on the differences of learning behavior and learning time in class and after class, with and without annotation sharing between HLA and LLA learners.
Learning behavior and learning time in class with/without sharing

In this section, a series of t-test analyses were conducted to examine the difference in learners’ learning behavior and learning time in class, with and without annotation sharing between the HLA and LLA learners.

No significant difference on learning behavior in class without annotation sharing was found between the HLA and LLA learners ($p > 0.05$). However, Table 8 revealed significant differences between the HLA and LLA learners on quantity of text annotation ($t = 1.789$, $p < 0.01$) in class with annotation sharing. This implies that the HLA learners’ frequencies of using text annotation in class with annotation-sharing mechanisms were significantly higher than those of the LLA learners.

<table>
<thead>
<tr>
<th>Table 8. Independent-sample t-test analysis in class with sharing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning achievement in class</td>
</tr>
<tr>
<td>--------------------------------</td>
</tr>
<tr>
<td>Quantity of text annotation in class</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Quantity of voice annotation in class</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

* $p < .05$. ** $p < .01$.  

Note.  

Figure 9 shows the line charts of different learners’ learning time in class. As shown in this figure, the HLA and LLA learners have different learning time by using the AME system with annotation sharing in class.

In order to examine the differences in learners’ learning time by using the AME system with/without annotation sharing in class between the HLA and LLA learners, analyses were undertaken between the HLA and LLA learners for the learning time. The result revealed that the difference level of learning achievement learners did not show any significant difference on their learning time by using AME system without annotation sharing in class ($p > 0.05$). However, there were significant differences in learners’ learning time by using the AME system with annotation sharing in class between the HLA and LLA learners ($t = 2.318$, $p < 0.05$), as shown in Table 9. These results indicated that the HLA learners were continuous learning in class, but the LLA learners didn’t.

<table>
<thead>
<tr>
<th>Table 9. Independent samples t-test analysis in class with sharing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning achievement</td>
</tr>
<tr>
<td>-----------------------</td>
</tr>
<tr>
<td>Learning time in class</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Learning behavior and learning time after class with and without sharing

In this section, we discuss a series of t-test analyses conducted to examine the difference in learners’ learning behavior and learning time after class with and without annotation sharing between the HLA and LLA learners. No significant difference was found between different levels of learning achievement learners on learning behavior after
class without annotation sharing ($p > 0.05$). However, the results from Table 10 show significant differences between the HLA and LLA learners on quantity of text annotation ($t = 3.017$, $p < 0.01$), voice annotation ($t = 2.625$, $p < 0.05$), and listening to one’s own voice annotation ($t = 2.662$, $p < 0.05$) after class with annotation sharing. This implies that the HLA learners’ frequencies of using text annotation, voice annotation, and listening to one’s own voice annotation after class with annotation-sharing mechanisms were significantly higher than those of the LLA learners.

The result of the learners’ interviews showed that their learning behaviors of using annotation-sharing tools are different. The HLA learners stated:
- “I would like to choose HLA students to review their annotation, because their annotations were usually great quality.”
- “I would like to choose students that made a lot of annotations.”

We found that the HLA learners preferred to use annotation-sharing tools to review learners who made more great annotations, because the HLA learners agreed that those reviewable annotations could help them learn. However, the LLA learners stated:
- “I would like to choose my friends’ annotation to review.”
- “I would like to arbitrarily choose other students to review their annotation, because I don’t know who to choose.”

We found that the LLA learners don’t prefer to use annotation-sharing tools to review HLA learners. In summary, there were differences in learning behavior that learners used annotation-sharing tools between HLA and LLA learners. This is why different learners’ after-class learning behaviors are different.

| Table 10. Independent-sample $t$-test analysis after class with sharing |
|-----------------------------|--------|------|---------|---|---|
|                             | Learning achievement | $N$  | Mean    | $SD$  | $t$  | Sig |
| Quantity of text annotation after class | High    | 15   | 2.21    | 2.31  | 3.017 | .008** |
|                             | Low     | 16   | 0.37    | 0.53  | 2.625 | .020*  |
| Quantity of voice annotation after class | High    | 15   | 0.75    | 1.01  | 2.662 | .019*  |
|                             | Low     | 16   | 0.06    | 0.12  |       |       |
| Quantity of listening one’s own voice annotation after class | High    | 15   | 1.01    | 1.43  | 2.662 | .019*  |
|                             | Low     | 16   | 0.03    | 0.05  |       |       |

The different level of learning-achievement learners’ learning time after class is shown in Figure 10. The HLA and LLA learners have different learning times by using the AME system with annotation sharing after class.

To examine the differences in learners’ learning time by using the AME system with or without annotation sharing after class between the HLA and LLA learners, we conducted a $t$-test analysis. The results from Table 11 show that the different level of learning achievement learners did not show any significant difference on their learning time by using the AME system without annotation sharing after class ($p > 0.05$). However, there were significant differences in learners’ learning time by using the AME system with annotation sharing after class between the HLA and LLA learners.
learners \( t = 2.619, p < 0.05 \). These results indicate that the HLA learners continuously learned after class, but the LLA learners didn’t.

<table>
<thead>
<tr>
<th>Learning achievement</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>15</td>
<td>8.26</td>
<td>9.17</td>
<td>2.619</td>
<td>.019*</td>
</tr>
<tr>
<td>Low</td>
<td>16</td>
<td>1.90</td>
<td>2.12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 11. Independent-sample t-test analysis after class with sharing

*Correlation between reviewing peers’ annotation and learning achievements with annotation sharing*

As shown in Table 12, the results of Pearson’s correlation analysis showed that the quantity of review HLA learners’ text annotation is related to learning achievement \( r = 0.413, p < 0.05 \). In addition, we also conducted a simple regression analysis to examine the learners’ learning behavior with an annotation-sharing mechanism as predictors for their learning achievement. The results demonstrate that the quantity of review HLA learners’ text annotation were the significant predictors for their learning achievement \( p < 0.05 \). It indicates that learning behavior in which learners review HLA learners’ text annotation can facilitate their learning achievement.

<table>
<thead>
<tr>
<th>Annotation sharing</th>
<th>High learning achievement</th>
<th>Low learning achievement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text annotation</td>
<td>.413</td>
<td>.252</td>
</tr>
<tr>
<td>Voice annotation</td>
<td>.289</td>
<td>−.049</td>
</tr>
</tbody>
</table>

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

### Implications and suggestions

Consistent with the results of previous research, our data reflect the advantages of using annotation for learning. For example, annotation for learning helps students recall teaching materials (Peverly et al., 2013), understand the content of the teaching materials more easily (Bohay, Blakely, Tamplin, & Radvansky, 2011), and focus on teachers’ explanations (Hwang et al., 2011b). This study also demonstrated that annotation-sharing mechanisms effectively enhance learning, which is consistent with the findings of Chen and Liu (2012), Su et al. (2010), and Yang, Tseng, Shin, and Liang (2012).

Furthermore, we suggest that learning behavior and learning achievement after class may be affected by parental input and supervision. Therefore, we recommend that parents are provided with the means to supervise and be involved in their children’s learning to enhance after-class learning. There is also a need for a suitable incentive for children to improve their learning after class. One example of this is the annotation-sharing mechanism, which can motivate pupils and help them to engage in active learning after class. Our results also indicate that the annotation-sharing mechanism has significant positive effects on learning achievement. There is, however, a need to provide guidance for peer-to-peer reviews of annotations. It is particularly important to provide learners with good annotations as references, thereby enhancing their learning.

### Conclusion

#### Summary of findings

Our major findings pertain to the following five key areas:

- **Learners’ perceptions.** The results from the questionnaire revealed that most learners had a positive experience with the AME learning system. Most students agreed that the multimedia tools provided by the e-readers helped them to learn English.

- **Parents’ perceptions.** Most parents believed that the AME system helped their children learn English. Additionally, most parents indicated that they would allow their children to use e-readers again as they believed...
the e-readers helped their children to study outside of the classroom. The learning behaviors observed by parents were significantly correlated with the learning behaviors outside of the classroom, as evidenced by the quantity of voice annotations and the time pupils spent listening to their own voice annotations. This indicates that parental involvement plays a key role in the frequency with which voice annotations are created and listened to after school. Students were more likely to practice outside of class under the guidance and support of their parents. Thus, the learning behaviors observed by parents were highly correlated with learning achievement.

- Learning behavior using the AME system. Learning behavior involving text annotations in class was significantly correlated with learning achievement. Additionally, the quantity of the text annotations made in class was a significant predictor of learning achievement. Therefore, effective text annotations in the classroom setting improved learning achievement.

- Differences in learning behavior and learning time in/after class. We found no significant difference between the in-class learning behavior of high-achieving learners (HLA) and that of low-achieving learners (LLA) when annotation sharing was not included in the analysis. However, significant differences between HLA and LLA learners regarding the quantity of text annotations made in class were observed when annotation sharing was included in the analysis. HLA learners used significantly more text annotations in class than did LLA learners. This suggests that HLA learners were better at using annotation-sharing mechanisms than were LLA learners, resulting in more efficient in-class use of this learning tool by the former.

Similarly, we found no significant differences in the learning behavior of HLA and LLA learners outside of class when annotation sharing was not included in the analysis. However, significant differences in the quantity of text annotations, voice annotations, and time spent listening to one’s own voice annotations after class were observed when annotation sharing was included in the analysis (i.e., these behaviors were more frequent in HLA learners). This again suggests that HLA learners may be better at using the annotation-sharing mechanism than are LLA learners, rendering the former more likely to use this tool outside of class.

We also found significant differences between HLA and LLA learners in the time spent learning using the AME system with annotation sharing both in and after class. HLA learners learned continuously, whether in or outside of class, whereas LLA learners did not. These results imply that the annotation-sharing mechanism is efficient and facilitates learning for HLA learners. This may be due to the fact that additional help is needed for LLA learners to learn English; for example, they may need to use the recommended tools and have access to the high-quality annotations made by HLA learners. This may help LLA learners learn English after class on their own.

- Learning behavior with annotation-sharing mechanisms in the AME system. A Pearson’s correlation analysis showed a relationship between learning achievement and how much or how often HLA learners reviewed their text annotations. This indicated that effective review of these annotations facilitated learning achievement for HLA learners.

Limitations

The present study has several limitations. First, only a small sample was used. Further work with a larger sample is needed. Another limitation is the method of calculating the amount of time spent learning; a more effective method would enable a more detailed analysis.

References


Appendix A

Parents' perception questionnaire

Parental perceived ease of use
PE1. By AME system, children can learn English easily.
PE2. By AME system, children can easily read the vocabulary.
PE3. By AME system, children can easily review content in class.
PE4. By AME system, children can easily have conversation by English.

Parental perceived usefulness
PEU1. Learning by AME system enhances children's English learning.
PEU2. Learning by AME system helps children repeatedly practise English words.
PEU3. Learning by AME system allows children to listen to teachers’ explanation of class at home.
PEU4. By writing assignments with AME system, children can practise dialogue in simple English at home.

Learning behavior observed by parents
LBP1. My children learn English by AME system at home.
LBP2. My children practise reading English words by AME system at home.
LBP3. My children review content of class by AME system at home.
LBP4. My children practise simple English conversation by AME system at home.

Appendix B

Learners' feedback questionnaire

Perceived ease of use
PE1. With the AME system, I can easily obtain system operation information needed.
PE2. It is not difficult to use AME system.
PE3. When using the AME system, the operation and interaction with the system are easy.

Perceived usefulness
PEU1. Use of the AME system enhances my English proficiency.
PEU2. Use of the AME system effectively enhances my English learning.
PEU3. Use of the AME system enhances my learning English.
PEU4. Use of the AME system allows me to have English learning rapidly.
PEU5. Use of the AME system enriches my annotation content.

Intention to use
IU1. I intend to accomplish English learning with the AME system.
IU2. In English learning, I frequently learn with the AME system.
IU3. Generally speaking, I am satisfied with learning English with the AME system.

Usefulness after school
EUS1. By using the AME system after school, I can learn English rapidly.
EUS2. Using the AME system after school reinforces my English proficiency.
EUS3. Using the AME system after school effectively enhances my English learning.
EUS4. Using the AME system after school helps my English learning.
EUS5. By using the AME system after school, I can learn English rapidly
EUS6. Using the AME system after school enriches my annotation content.
EUS7. Using the AME system after school considerably enhances my English learning.

Intention to use after school
IUS1. I intend to learn English with the AME system after school.
IUS2. I intend to accomplish English learning activities with the AME system after school.
IUS3. In English learning after school, I frequently learn with the AME system.
IUS4. Generally speaking, I am satisfied with English learning with the AME system after school.
IUS5. If possible, I will learn English with the AME system after school.
Mobile Enhanced Learning in a South African Context

Mmaki Jantjies¹ and Mike Joy²

¹Department of Information Systems, North West University, Mafeking, South Africa // ²Department of Computer Science, University of Warwick, Coventry, UK // mmaki.jantjies@nwu.ac.za // m.s.joy@warwick.ac.uk

*Corresponding author

(Submitted August 31, 2013; Revised March 29, 2014; Accepted June 7, 2014)

ABSTRACT

Multilingual classrooms in developing countries are often challenged by a lack of digital resources and technology which supports their multilingual learning process. Code-switching is a phenomenon common to multilingual schools where learners are taught in a language which is not their first language. In these environments, code-switchers frequently alternate between their first and second languages, seeking alternative words to clarify their understanding of the topic being studied. This paper presents a study based in South Africa, where mathematics learners interacted with a mobile learning system named M-Thuto, supporting learners who code-switch while learning and providing them with digital resources. The system consisted of summarised notes, class exercises and a class quiz. Through semi-structured interviews and questionnaires, data were gathered and analysed from 90 learners to gain perspectives on their interaction processes. The study aimed to establish how mobile learning can be used to support multilingual learners in under-resourced schools. The results of the study reflect the need for mobile learning resources that support their learning considering their linguistic challenges. The results also reflect the important role that mobile phones can play as alternative digital learning resources.

Keywords
Multilingual, Code-switching, Mobile learning, Setswana, South Africa

Introduction

In many developing countries education is often provided in a learner’s second or third language, and thus learners sometimes struggle to appropriately interpret tasks and content, as a result of inadequate language acquisition in the language of learning and teaching (LOLT) (Arthur, 1996; Setati and Planas, 2012). While there has been extensive research into mobile learning for second language acquisition (Collins, 2005; Ballance, 2012; Wong et al., 2012) there is limited research that focuses on ubiquitous learning that can be used to support learning in multilingual environments.

Language forms an important part of the learning process as the learning language not only affects the learner’s communication within their school but it also affects their interpretation and understanding of the learning content, especially in subjects such as mathematics which require proficient knowledge of the LOLT (Abedi & Lord, 2001).

South Africa is a multilingual country with 11 official languages. Even though there is officially one LOLT, the South African government supports the use of other languages to assist the learning process in schools (DOE, 1997) as the majority of the country’s population are not English first language speakers. The South African education system is challenged by multilingual classrooms with limited learning material to support learning in these types of environments (Maree et al., 2006). With eleven official languages, teachers often face the challenge of teaching learners from multilingual backgrounds through a language that they only accurately acquire when they begin formal education. This often leads to continuous code-switching within classrooms. Code-switching is a phenomenon common in multilingual environments where a speaker always reverts to the use of two languages in a conversation. The reasons for alternating the languages could be as a result of an inadequate language vocabulary in their second language and thus the speaker will always go back to their mother tongue to seek the relevant duplicate word (du Plessis & Louw, 2008). This problem is not unique to South Africa and is mostly observed in learning environments where the LOLT differs from the learner’s home languages (Lin, 1996; Arthur, 1996; Moschkovich, 2007).

Developing mobile learning applications that can be used in subjects such as physical sciences, mathematics, history, etc., to support such environments requires appropriate content which considers the user’s challenges and needs.

The objective of the research project reported here was to establish potential solutions to the following research questions.
What are the challenges associated with learning content that code-switching learners face in formal learning environments?

What are the resource needs of learners in under-resourced learning environments?

How can mobile learning be used to support learners who code-switch in under-resourced environments?

This paper evaluates, through a case study, a proposed mathematics bilingual mobile learning application named M-Thuto which was developed to support learners in South Africa. M-Thuto is a web-based tool which is accessible on a WAP enabled mobile phone, and aims to be a comprehensive learning tool consisting of the learner’s class-notes, interactive class exercises with suggested worked solutions, and a class quiz to test the learner’s understanding of the learning content. M-Thuto uses the code-switching technique of learners to present mathematics content in both the Setswana and English languages. The study was conducted in a predominantly Setswana speaking area, while English is the LOLT in the schools under study. The name M-Thuto is derived from both the English and Setswana language which is mobile education (thuto). The development of this application was also motivated by the shortage of reading and learning material in some South African high schools.

The study presented in the paper was facilitated over a period of a month with learners from four different schools interacting with the application and subsequently giving individual feedback at the end of the study. Subsequent to the interactions, surveys were conducted with learners giving feedback on the learning process and the language support of M-Thuto on their mathematics subject. Teachers also participated in the study with the M-Thuto system giving them provision to monitor the learners’ interactions with the tool through the teacher’s personal access portal. The results of the study provide support for methods to be considered when creating a multilingual mobile learning environment towards a proposed multilingual mobile learning framework.

Literature background

Mobile learning

Mobile learning has been defined by different researchers, with each of their definitions being influenced by their perspectives on the role it plays in the learning environment. Some mobile learning definitions follow a more technical approach which concentrates on the mobile learning platform, while other definitions focus on the pedagogy experience and the ubiquitous nature of the mobile learning process (Traxler, 2007). The working definition used in this paper will proceed from the description provided by Kukulska-Hulme et al. (2009, p. 20) that the process of mobile learning is affected by ubiquitous learning practices that, “cross spatial, temporal and/or conceptual borders and involve interactions with fixed technologies as well as mobile devices. Weaving the interactions with mobile technology into the fabric of pedagogical interaction that develops around them becomes the focus of attention.” In this paper we consider the learning language as an important part of the pedagogical interaction process. Mobile learning provides a learning platform across learning spaces through the contextual language as a learning resource in which pedagogy is delivered through the ubiquitous learning process.

Multilingual learning content

The Internet is one of the first and most widespread methods of education delivery through electronic media, with people accessing it on their mobile and desktop devices. Until recently, English was generally the primary language used to deliver online content. The use of Arabic and Chinese languages on the Internet has grown rapidly in recent years, and this has led to an increase in the number of indigenous language speakers using the web. The translation of web pages into other languages has also reflected a proportional growth in Asian, Arabic and European languages (Chung, 2008), and there has been a consequent increase in the number of multilingual websites. Previously there was an absence of African languages used in electronic media, with no options provided in the language choice features of any form of electronic tool (Osborn, 2006). Even though there has been progress in the use of African languages in electronic media, African language options still remain limited.

In mobile learning empirical studies, there is evidence of the effectiveness of mobile learning applications which aim to support content accessibility in multiple languages across different subject areas. For example, Yoza is a mobile novel book application that was initiated to mitigate the lack of access to books by South African youth. The
application provides mobile device accessible novels which are available in both English and Xhosa (one of the 11 official South African languages). The content of the application is viewed on a mobile phone through a Mixt platform which provides online access to different resources at affordable prices. The Yoza initiative has grown to include access to broader literature which includes short stories, novels and poetry (UNESCO, 2012). After each session the readers can give feedback on the story they have read. UFractions is another example of an m-learning tool replicated in multiple languages which is available in English, Finnish and Portuguese, and has been piloted in South Africa, Finland and Mozambique. UFractions is a mathematics mobile game that enables high school learners to learn fractions by solving some tasks on their mobile phones (Turttaiinen et al., 2009; Laine, 2011). Each of these applications was developed for a specific context. Despite the context and language support provided by each of these m-learning applications, code-switching features and language adaptation elements were not utilized to fully support the bilingual learner.

The proposed multilingual mobile learning content design

Multilingual learning environments and the practice of code-switching

Code-switching is defined as the use of more than one language in an utterance or during a conversation. The behaviours of code-switchers often differ between the informal and formal contexts. In formal learning environments, especially in formal writing, code-switching occurs less frequently compared to informal environments (Graedler, 1999). While other researchers view code-switching as a barrier to mastering the LOLT in classrooms (Pfaff, 1979), most researchers view code-switching as a unique strength that second language speakers of LOLT have to view and understand learning in more than one language (Arthur, 1996). In learning environments, code-switching is used as a scaffolding mechanism that can help learners who are not first language speakers of the instructional language to understand learning content in a second language (Adler, 1998; Fennema-Bloom, 2009). In a typical classroom where code-switching occurs, the teacher will move between the language of instruction and the learner’s home languages to either reinforce the learner’s understanding or to introduce new challenging topics (Setati, 1998; Fennema-Bloom, 2009).

The effect of code-switching is often more visible in classrooms where the teacher is also not a first language speaker of the LOLT (Ncoko et al., 2000; Setati & Adler, 2000; Probyn, 2006; Moschkovich, 2007; Carless, 2004; Mbali & Douglas, 2012). In such environments both the teachers and learners always revert back to code-switching and attempt to recreate the content in the learner’s first language, to give a clear understanding of the learning content. Setati (1998) observed a classroom in South Africa to establish how code-switching occurred. In her studies she revealed that teachers in formal classrooms use code-switching in the following ways.

- To reinforce a concept – this occurred when a learner would ask a question in class related to what the teacher had just taught using the official LOLT. The teacher would revert between the learner’s home language and the LOLT to give clarity to the learner’s question.
- To explain new concepts – this occurred when a teacher was introducing a new concept in class. The teacher would explain the topic in the LOLT and would then explain the topic in the learner’s home language.
- Translation – this occurred when a teacher would directly translate text from the LOLT to the learner’s home language.

In this paper we consider Setati’s (1998) code-switching methods and use them to create learning content that can be used to support mathematics learners through the M-Thuto mobile learning environment.

Case study

Learners from different schools based in both urban and rural locations participated in this case study. The difference in locations outlined the difference in learners’ needs and challenges related to language and access to electronic learning resources. In the study, learners were provided with mobile phones during their mathematics classes and required to interact with the system and provide feedback on their interaction process. After using the system, learners provided feedback through interviews and questionnaires. The languages used for the M-Thuto system were English and Setswana, as Setswana was the common native language of the geographical location were the study was conducted and the English language was the language of teaching and learning in the participating schools.
The M-Thuto software

The architecture of the application in this study is based on a client server layout. The learning content is accessible through any WAP enabled mobile phone. The learner’s mobile phones access the application and related content on the server, and an SQL database is also stored on the server. The database includes the learner’s interaction details and results from their quiz. The choice of this architecture was influenced by the available network and resources, and it was important to develop content that would be accessible even on the most basic web-enabled mobile phones. Each learner has their own profile keeping records of their individual interactions and performance. Basic open source web development languages XHTML and PHP were used to design the application interface and enable communication with the database during learner interaction. The learning content was developed in consultation with a mathematics specialist who was previously a teacher and now a head of department in the local government where the research occurred. Teachers and language translation specialists also participated in the content development process.

The purpose of the application was to provide a comprehensive tool that also reduces the problem of resource shortage faced by some of the participating schools. The learning content was structured in a menu using the as follows.

Class Notes – The first section of the system was a notes section, these notes were simplified teachers’ notes of the topic of simultaneous equations that they had covered in class.

Class Exercises – The second section was a class exercise questions page that allowed learners to practice potential examination questions. Each question had a hidden answer that a learner could only reveal after attempting the question. The learners would then view the correct answer to the questions reflecting on other methods they could have used to answer it.

Class Quiz – The third section was a class quiz which allowed learners an opportunity to test their understanding of the topic area. Their answers were sent to the database for teachers to keep track of.

Attributes – The final section was as an acknowledgement section that acknowledged the parties that contributed to the development of the learning content including the translation of the content.

The M-Thuto activity menu choice

![Activity Menu Choice](image)

*Figure 1. The learner interaction pages with M-Thuto (Jantjies and Joy, 2014)*


**Code-switching attributes**

The content support for code-switching of this system was based on Setati’s (1998) code-switching type principles. Setati mentions three ways in which code-switching is used in the classroom learning environment. These three approaches are adapted to the mobile learning environment to present three methods that can be used to create a mobile learning system to support the code-switching behaviour.

*Introducing content activity:* When new concepts are introduced to learners, learning content is provided to learners in micro-paragraphs, using the LOLT and then the same content is recreated in a different language using the learner’s home language and presented to the learner. Appropriate navigation features are required as the learners will need to move between the learning pages to achieve this code-switching support.

An example of introducing content activity in the M-Thuto system:

![Image](image.png)

*(Jantjies and Joy, 2014)*

*Reformulation:* In this form of code-switching learners are provided with a rephrased wording of the same content in a different language. This support is provided with lines of content in different languages preceding each other.

An example of reformulation in the M-Thuto mobile quiz:

![Image](image.png)

*(Jantjies and Joy, 2014)*
Translation: this is the direct translation of words, but is often not used in classrooms to facilitate the code-switching environment. This is as a result of direct translation often distorting the meaning of the content. In this paper we suggest an alternative means of translation by giving learners an option to view an alternative word in another language. A similar feature has been used by the North West University (Van Huyssteen et al., 2007) to support multilingual learners. This paper proposes a selection option of the word within text which enables the learner to view an alternative word in another language (the learners home language) giving a direct translation of the word. On the mobile application a learner would be given a text and select a word which they are not clear to view its alternative in another language instead of viewing the full paragraph in another language as presented above in the Introducing content activity example.

An example of translation:

![Translation Table]

The sun brings heat

(Jantjies and Joy, 2014)

The research methodology

A sample of 90 learners from four public schools within the North West province participated in this research. The schools were situated in different geographic setting with one school situated in a township, two schools located in rural villages and the last school in an urban area. The learners’ ages ranged between 16 and 18 years with these learners distributed across grades 11 and 12, which are the last two years of secondary school in the South Africa education system. All learners in the research process were registered for mathematics as one of their school subjects. In order to conform to ethical requirements, a formal request to perform research was submitted to the North West department of education.

Mobile phones are prohibited for use in schools and a request was made to each school to allow their use for the research, and the learners each received a WAP enabled mobile phone for the research process. Despite this, some learners came with their own mobile phones and interacted with the tool through their own phones. The tool was accessible online with each mobile phone having airtime to access the Internet. Some lessons during the week were dedicated to this mobile learning process. Each individual learner chose an activity that they wanted to focus on for the day.

Content developed for the research process was based on the topic area of simultaneous equations with topics ranging from grade 10 to grade 12 content. Each page of content had an interchangeable option with a choice to view it either in Setswana or English, allowing the learner to change their view at any time during their study. The content was divided into three learning areas: notes designed by teachers for daily learning, a section consisting of class exercise questions and answers in the form of a drilling method, and finally a section to test learners’ understanding of the topic through different tests. The learners spent most of their time on the class exercises as they enjoyed receiving real time guided alternative answers to the questions they had attempted during the interactive drill and answer process.

Learners subsequently filled in questionnaires on their interaction process with the researchers present, allowing the learners an opportunity to ask for any further clarity on the questionnaires. The questionnaires allowed the respondents to be free in expressing their views without the feeling that they were being monitored, and mostly consisted of closed questions. Learners were also interviewed through semi-structured interviews, which are guided but are flexible in allowing the interviewer to gain more content beyond the guided question format (Cohen et al., 2011). The research took place over a month.
Data were divided and discussed through the thematic analysis process which concentrates on the response of the participant as opposed to the manner in which they responded (Bryman, 2004). The data gathered from the interviews and questionnaires are discussed below, with fictitious character names to hide the participants’ identities.

Study results

Code-switching challenges

This section was based on Moschkovich (2010) who explains that in order to understand a multilingual learning environment the researcher needs to know the linguistic background of the learners and how their bilingual nature affects their learning process. This section sought to establish the linguistic backgrounds of learners and how being bilingual affected their learning process in their mathematics class.

The types of questions in the surveys included the following:
- How many of the South African languages can you speak, read and write?
- Do you ever have challenges in understanding content when provided with content or instructions through the language of learning and teaching which is English?

Results

100% of the learners were able to communicate effectively in two or more languages which included their home language and English which was the LOLT. From the eleven official languages, 89% of the participants could effectively communicate in Setswana and English. The high number of first language Setswana speakers was influenced by the geographic location of the study. Knowing the languages specific to a geographic location enables content developers to create content which considers the specific locations giving users language options familiar to them. When asked about their challenges in understanding the language content presented for monolingual speakers, a collective of 53% of participants agreed that they often incurred problems understanding content as a result of the language that it is presented in. As seen in figure 2, when looking further at the data, the schools reflected the difference in the local area use of languages. Schools located in the rural areas and townships where English was not commonly used for communication, had the highest number of children consenting to this challenge with 67% of learners in schools C and D, followed by 60% in school B. School A, located in an urban area where English was commonly used for communication along with Setswana, presented only 42% of learners as having this problem while learning. Learners from urban schools such as school A, had also often received good English foundation in their primary school years. From these learners, a collective of 63% of the learners stated that thus often code-switched while learning during class.

![Figure 2](image_url)

*Figure 2. The percentage of learners who struggle to understand content*
The role of mobile learning in an under-resourced environment

As schools in low income areas would also struggle with having sufficient learning resources to support each learner, the questions in this section sought to establish the electronic learning resources that participating learners already had access to which could give them access to additional learning material. The questions also sought to establish if the existing digital resources made provision for their bilingual learning. Considering the low cost of mobile phones and high access that learners have to mobile phones this section aimed to establish the role that mobile learning could support the learning process in under-resourced bilingual environments.

The types of questions in the surveys included the following.

- Do you own a mobile phone?
- If your answer for the previous question was no, do you have access to a computer at home?
- Are you ever needed to use electronic devices as part of your learning (i.e., search the Internet to gather data?)

Results

From the four participating schools only one school had computers available for learning within the school. All participating learners either owned or had access to a mobile phone at home, while there was poor ownership and access to computers. The responses however differed with each school location with school A in a high income area displaying a different array of results from learners in schools B, C, and D which were based in low income areas.

When questioned about their history with electronic learning, learners from school A mentioned that they used the Internet to access learning resources and knew where to access further learning resources through their mobile phones. The teacher in school A was also very knowledgeable about external resources that learners could use to supplement their learning process and motivated them to use those resources but in their informal home environments.

The learners in the low income areas (schools, B, C, and D) were completely reliant on paper based learning material received from their schools and could not afford learning material beyond this. Many of these learners were unaware of web-based or electronic and mobile learning resources and games.

From the participants in all schools, 56% of learners mentioned that they were sometimes given tasks that required the use of technology through either online research or typing their work. However, as seen on figure 3, only 22% of participating learners could access a computer to perform these tasks, whereas 48% of the participating learners owned a mobile handset and other learners had access to them. Access to mobile phones in comparison with other electronic learning devices reflected the potential that mobile learning had in supporting the learning process of the participants. All learners had access to a mobile phone at home. Mobile learning also had a potential for providing more learners with access to resources as most of the participating learners did not have access to other digital learning resources and were at times required to use computers for doing their homework.

![Figure 3. The percentage of all participants' access to technology](image-url)
In schools C and D learners would at times not have enough learning material such as prescribed text books for each learner to use at their own convenience. This meant that learners would occasionally have to share learning material which they would have access to in order to obtain learning material. Having access to mobile phones, provided them with access to mobile Internet which served as an additional resource to accessing learning material such as online mobile books.

**Using mobile learning to support learners who code-switched in under-resourced environments**

The questions in this section sought to establish the learner’s previous experiences of electronic learning and their experiences in using the M-Thuto system. The objective of this section was to establish the applied use of content view principles in supporting the code-switching learners through a mobile learning system and how this system could support schools challenged by a lack of electronic learning resources.

The types of questions in the questionnaire included the following.
- Have you ever used a mobile application for learning?
- If your answer for the previous question was yes, was the application a learning system or an educational game?
- In the M-Thuto system which of these languages did you use to read the learning content?
- Do you find the system effective in supporting the learning process?
- What are the challenges you faced with the mobile learning process?

**Results**

From the participating learners, 53% of them had previous experiences of interacting with mobile education applications, but all of these interactions were only game based and were not directly related to their current studies. All of the learners with previous experience came from the urban school A, which is consistent with the results above that learners from schools B, C, and D had limited access and knowledge of electronic learning resources. M-thuto was used to support learning during lessons and thus compensating for the resources such as text books which were needed by each learner as the system was able to provide notes and exercises which could also be accessible through their books. During this study every learner thus had access to the resources at their own convenience as the system also gave them access to the Internet to enable them to use online mathematics resources.

From the interaction process with M-Thuto, as presented in figure 4, 61% of all learners had used both English and Setswana to read the content of the application. 34% of the learners used only the English language to read content while 98% of all participants mentioned the effectiveness of the bilingual learning tool in supporting their learning process. Looking at this data further the schools in the rural and under-developed areas had the highest number of viewings in using both Setswana and English to read the content. At the end of the study the learners and teachers felt that the project should go beyond a pilot study as the learners found the mobile learning interaction process helpful towards achieving their learning objectives.

![Figure 4. The percentage of all participants’ view of content using English and Setswana](image-url)
School A in the urban area had the highest response percentage of learners using only English to read the content. School C, which is located in a rural village, had the highest number of learners viewing content in both languages. Not a single learner used only Setswana to view the content.

Despite the code-switching nature of learners throughout the schools, the ability to read their home language was not common to all learners. The learners in the rural schools seemed to appreciate the language aspect more than those who came from the urban school. From school D, Basadi expressed the experience as “fun because we were learning on mobile phones and this made me look forward to learning. Reading content in both Setswana and English also made the process exciting because I explored my Setswana.” Even though the majority of learners in schools A and B used only English to view content, most of the original language views of learners in these schools had become less stringent after their mobile learning experience. The learners were now more focused on the look and feel of the learning tool and the need to expand the tool into other subjects, giving suggestions on how to make the application interaction more fun. From school A, Itumeleng felt, “it is nice to be able to move back and forth between the languages when reading the content. If one of the languages gets too difficult I can switch back to check what the term means in another language. I however feel that the application needs more ‘flash’ and sound.” Most of the learners in school A attended the ICT lessons which are offered in the school. Also from school A, Lesego said that “I loved the tool but I just feel that there is too much focus on mathematics, other subjects that we find difficult are neglected. Is it possible to use this tool for my other subjects?” In the rural schools where learners were very opinionated about using their home languages for learning, they took time out to read the content and learn using both languages.

**Discussion**

In the study above, we observed that some of the participating schools were still challenged by limited learning resources available to the learners. Even though school A (which is located in a high income urban area) was well resourced, the participating learners felt that ubiquitous learning resources provided further learning material which was effective in supporting their learning process. Learners from the remaining schools saw ubiquitous resources as a tool that would reduce their existing resource gap in support of their learning. Accessing further learning resources which they would not have to pay additional money to obtain, could enhance their learning as most of them cannot afford additional learning material (Legotlo et al., 2002).

As some schools had a challenge of providing learners with required reading material to support learning, we found that M-Thuto was able to support the need for this resource role by providing them with class notes and exercises which could also be found in their prescribed text books. Many parents in rural and low-income communities are often without a job or are employed in low wage roles such as housekeeping and gardening. These parents cannot provide beyond their means as their income is often only enough for the family basic living costs. Further costs related to their children’s education or school activities become difficult to support. Propelling this situation in low income areas are the poor resources, limited books and teacher shortage experienced by learners especially in vital learning areas that require efficient resources in order for a learner to perform well (DOBE, 2011). These schools are often out of reach of public amenities (i.e., libraries and community computers) that can provide free learning resources. Considering the wide access to mobile phones in all South African communities, mobile learning emerges as a potential solution that could be applied to reduce the resource shortages in these communities (Vosloo & Botha, 2009). Learners from low-income communities located far from cities also have limited knowledge of free online and offline resources that can support their learning process. Through access to basic technology such as mobile phones, learners can enhance their learning process by accessing these resources to support their learning.

The South African education system is also facing challenges of language use in schools with learners coming from different linguistic backgrounds. Learners who participated in this research found that code-switching in their learning process and amongst their peers helpful in supporting them to clearly articulate themselves and understand tasks. M-Thuto gave them an opportunity to alternate through different ways to their home language during their learning process. The problem of not understanding tasks or being unable to articulate responses without code-switching while learning seemed to be prevalent in mathematics lessons. Mathematics requires a clear understanding of the language that it is imparted in, as this could affect the learner’s interpretation of the task and their response to the task. The challenge of poor English language backgrounds and the unavailability of bilingual learning material contribute to the poor performances of learners in key learning areas such as mathematics (Botes & Mji, 2010; Setati, 2008).
The learners who had a previous experience of interacting with some form of mobile learning tool expressed that there was a significant absence of mobile learning content written in South African languages. There is a clear need to further explore the availability of mobile pedagogy delivered through various languages. Having access to appropriate mobile learning content available in a choice of languages can support learners in South Africa to enhance their learning process.

Conclusion

This paper presented a study where learners from four uniquely resourced schools interacted with mobile learning system named M-Thuto and presented their views on their interaction process. The research results from the case study, on the M-Thuto experience of the high school learners from both rural and urban school results, show that there are still learners who are in need of supplementary ubiquitous learning resources and materials. M-Thuto was able to provide them with learning notes, drill exercises and related quizzes to support their learning. Even though learners in urban participating schools could acquire and access supplementary learning material, having accessible and ubiquitous multilingual learning material assisted them to learn in their own time, in their choice of language, while introducing them to new methods of ubiquitous learning. Furthermore, by providing bilingual content support, M-Thuto gave learners the opportunity to have options in methods of viewing mobile learning content supporting their code-switching nature. Techniques such as content adaptation and efficient responsive memory provided effective design principles for developing the architecture to suit this context.

In a country where the majority of the learners are bilingual, learners must be afforded the opportunity to be able to choose which language would aid them to understand their learning content. The future of the M-Thuto system is to further enhance its architecture to support code-switchers, allowing them to access a context aware learning resource. Through various enhancements of the system a framework will be formulated which can be used to support the development of mobile learning software to be used in similar multilingual learning environments.

Mobile learning supporting bilingual or multilingual content needs to be further investigated to find effective ways to improve South African education and other under-resourced multilingual learning communities through appropriate and affordable technology.

Acknowledgements

This paper is a revised and extended version of the paper “Multilingual Mobile Learning – A case study of four South African High Schools”, published in Proceedings of the 11th World Conference on Mobile and Contextual Learning (Jantjies & Joy, 2012).

We would like to thank the South African North West Department of Education (DOE) for allowing us to perform research in the participating schools as well as thank the participating teachers and learners. We would also like to thank the North West DOE mathematics head of department and one of their Setswana department members for their contribution to the content design process. We would like to also acknowledge the South African National Research foundation for their financial support on this research.

References


Measuring Knowledge Elaboration Based on a Computer-Assisted Knowledge Map Analytical Approach to Collaborative Learning

Lanqin Zheng¹, Ronghuai Huang¹, Gwo-Jen Hwang² and Kaicheng Yang¹

¹School of Educational Technology, Faculty of Education, Beijing Normal University, Beijing // ²Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taipei // 
bnuzhenglg@bnu.edu.cn // huangrh@bnu.edu.cn // gjhwang.academic@gmail.com // yangkc_beijing@bnu.edu.cn

*Corresponding author

(Submitted February 10, 2014; Revised May 18, 2014; Accepted May 25, 2014)

ABSTRACT

The purpose of this study is to quantitatively measure the level of knowledge elaboration and explore the relationships between prior knowledge of a group, group performance, and knowledge elaboration in collaborative learning. Two experiments were conducted to investigate the level of knowledge elaboration. The collaborative learning objective in the first experiment concerned the understanding of curriculum objectives, and that of the second experiment was related to the theory and application of consumer behaviour in microeconomics. A total of 91 undergraduate students participated in the first experiment and 94 participated in the second experiment. Students were randomly divided into 30 groups of three or four in each experiment. Students’ interactions were analysed based on the computer-assisted knowledge map analytical approach to measuring the level of knowledge elaboration. Empirical evidence from 60 groups demonstrates that the network structure entropy, degree distribution index, depth, and weighted path length of the activation spanning tree of the target knowledge map can be used for the precise measurement of knowledge elaboration. The results also reveal that knowledge elaboration is positively related to both prior knowledge of a group and group performance.

Keywords

Knowledge elaboration, Collaborative learning, Knowledge map, Computer-assisted instructions

Introduction

It has been widely acknowledged that collaborative learning facilitates knowledge gains (Dillenbourg, 1999; Stahl, 2011). Moreover, elaboration is an important activity for promoting knowledge acquisition during collaborative learning activities (Denessen et al., 2008; Golanics & Nussbaum, 2008; Stegmann et al., 2012). Knowledge elaboration has been defined as organising, restructuring, interconnecting, and integrating new information with prior knowledge (Reigeluth et al., 1980; Weinstein & Mayer, 1986; Kalyuga, 2009). It can facilitate the retention of the target information (Anderson, 1983) and stimulate the reorganisation of information. Educators have pointed out that elaboration processes are necessary for meaningful learning, which emphasises the integration of new knowledge into existing knowledge (Novak, 2002; Kalyuga, 2009). The importance of elaboration is also supported by the generative model of learning put forward by Wittrock (1989), who indicated that new information should be meaningfully related to prior knowledge to generate connections between the informing information and memory representations in order to retain new information.

Researchers have further indicated that knowledge elaboration is the key factor in collaborative learning (Denessen et al., 2008). Many studies have reported that knowledge elaboration has positive effects on students’ learning achievements (Van Boxtel et al., 2000; Stark et al., 2002; Hwang et al., 2007; Denessen et al., 2008; Stegmann et al., 2012). However, most previous studies measured knowledge elaboration by questionnaire (Draskovic et al., 2004), by coding think-aloud protocols (Stegmann et al., 2012), by coding discussion transcripts into different categories (Eysink & de Jong, 2012), or by assigning values of -1, 0, and 1 (Ding et al., 2011) to assess knowledge elaboration. Such coding is subjective and ignores the domain knowledge when segmenting and coding discourse data (Suthers et al., 2010; Zheng et al., 2012). Little research has been performed to determine how to measure knowledge elaboration accurately and objectively.

The present study attempts to overcome the methodological limitations in measuring knowledge elaboration. The purpose of this study is to go beyond the scope of previous studies by measuring the level of knowledge elaboration based on the knowledge map analytic approach, and by examining the relationships between the prior knowledge of
a group, group performance, and knowledge elaboration. The knowledge map approach is quite different from concept mapping in terms of knowledge representation. The nodes in a knowledge map can denote symbols, concepts, principles and formulas, processes and steps, cognitive strategies, and facts and instances (Yang, 2010), while those in a concept map mainly denote concepts, facts and instances (Novak & Cañas, 2006). More specially, concepts are defined as abstract objects (Laurence & Margolis, 1999), or perceived regularities or patterns in events or objects (Novak & Cañas, 2006). Based on the knowledge map approach, the following research questions are investigated in this study:

- What are the indicators of knowledge elaboration from the perspective of graph theory and knowledge semantic properties?
- Does students’ knowledge elaboration level positively affect their group performance?
- Is students’ knowledge elaboration level positively related to their prior knowledge?

**Literature review**

Knowledge elaboration plays a crucial role in collaborative learning. When group members interact with one another, they need to explain large quantities of information about the learning material to others and therefore process information more deeply. Slavin et al. (2003) believed that elaboration can be achieved by explaining information to others when interacting during collaborative learning. Researchers assume that interaction with others promotes the processing of information and the modification of cognitive structures (Baker, 2003; Wibeck et al., 2007; Mitnik et al., 2009; Suthers et al., 2010). Therefore, interactions in collaborative learning can stimulate knowledge elaboration and consequently promote individual knowledge gains.

What is still unclear is the nature of knowledge elaboration in collaborative learning. Insights into measuring knowledge elaboration are an important step towards unravelling the nature of knowledge elaboration. However, there is no consensus regarding how to measure the level of knowledge elaboration in collaborative learning. De Leng et al. (2009) investigated elaboration by analysing the frequency and duration of activities related to inquiry, interpretation, and reflection recorded in log files and administering a questionnaire to assess students’ perceptions of the working process. Ding et al. (2011) endowed each message with an elaboration value:

-1, 0, or 1. Many researchers have developed coding schemes to analyse discourse data to measure knowledge elaboration. In Table 1, different instruments for measuring knowledge elaboration via a coding scheme are presented.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Analytical focus</th>
<th>Coding scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Van Boxtel et al. (2000)</td>
<td>Elaborative episodes</td>
<td>giving elaborated answers</td>
</tr>
<tr>
<td></td>
<td></td>
<td>elaborated conflict</td>
</tr>
<tr>
<td></td>
<td></td>
<td>reasoning</td>
</tr>
<tr>
<td></td>
<td></td>
<td>cognitive example elaboration</td>
</tr>
<tr>
<td>Stark (2002)</td>
<td>Example elaboration</td>
<td>meta-cognitive example elaboration</td>
</tr>
<tr>
<td></td>
<td></td>
<td>other elaboration</td>
</tr>
<tr>
<td></td>
<td></td>
<td>instrumental help seeking</td>
</tr>
<tr>
<td></td>
<td></td>
<td>help given with labeled explanations</td>
</tr>
<tr>
<td>Denessen et al. (2008)</td>
<td>Cognitive elaboration</td>
<td>challenging help received with labeled explanations</td>
</tr>
<tr>
<td></td>
<td></td>
<td>acknowledging help with labeled explanations</td>
</tr>
<tr>
<td></td>
<td></td>
<td>self-questioning with labeled explanations</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1: information was completely ignored by all four group members</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2: one of the members mentioned a crucial item of information, but no one reacted to it</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3: one of the members mentioned an item of information and at least one of the members reacted to it</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4: one crucial piece of information was mentioned by at least one member and at least two of the other three members reacted to it</td>
</tr>
<tr>
<td>van Ginkel &amp; van Knippenberg (2008)</td>
<td>Information elaboration</td>
<td>5: one crucial piece of information was fully discussed by at least three members and integrated with other</td>
</tr>
</tbody>
</table>
information

- 6: at least two crucial pieces of information were fully discussed by at least three group members and integrated with other information
- 7: all three crucial pieces of information were clearly and fully discussed by at least three of the four members
- Providing explanations
- Asking questions
- Acceptance with further elaboration
- -1: the message was off task and distracted the students’ attention

Prinsen et al. (2009) Elaborated contribution

Ding et al. (2011) Knowledge elaboration values

- 0: the message was a task-related message but did not improve the solving process
- 1: the message was pertinent to the task and contributed to the final success of the problem solving
- Developing and testing hypotheses


- Relating and integrating
- Providing (self-) explanations
- The depth of cognitive elaboration can be computed by the duration of cognitive elaboration (multiply the number of segments by the length of the segments) according to think-aloud protocols

Stegmann et al. (2012) Depth of cognitive elaboration and argumentation-related cognitive elaboration

- The transcripts of argumentation-related cognitive elaboration was coded into claim, grounds, and qualification according to think-aloud protocols

There are some limitations to using these coding schemes to measure knowledge elaboration. Firstly, the coding schemes only focus on speech acts in discourse transcripts, such as providing explanations, asking questions, or acceptance with further elaboration. In fact, judging which messages belong to which types of speech act is both ambiguous and subjective (Strijbos, 2006). In addition, coding is very difficult because the purposes of humans’ speech acts are implicit; thus, the identification of speech acts is also subjective and lacks a reference point (Zheng et al., 2012). Secondly, these coding schemes ignore domain knowledge construction, which runs counter to the definition of knowledge elaboration. Thirdly, assigning each speech act to an isolated meaning does not record the indexicality of the meaning (Suthers et al. 2010). Therefore, it is necessary to develop a new method to precisely quantify knowledge elaboration.

Since the internal knowledge structures during the knowledge elaboration process cannot be directly observed, we seek to externalize them based on a graph theory approach. Previous studies have reported that graph theory is a promising educational diagnostic approach (Ifenthaler, 2010; Pirnay-Dummer et al., 2010). Furthermore, representing knowledge and relationships between knowledge in graphs has been recognized as being an effective way of evaluating students’ knowledge structure as well as helping them organize knowledge (Hwang, Yang, & Wang, 2013). Knowledge elaboration has been conceptualised as processes that connect new knowledge with existing knowledge structures (Weinstein & Mayer 1986). In this study, some measures from graph theory are employed to diagnose the level of knowledge elaboration.

Methodology

Development of knowledge elaboration indicators and the computer-assisted analytic tool

Indicators of measuring knowledge elaboration levels

In order to measure the level of knowledge elaboration, we designed a set of graphical indices to represent the structural and semantic properties of target knowledge maps, which are composed of the target knowledge selected and identified by designers according to the collaborative learning objectives. We hypothesize that the following attributes of knowledge maps can serve as indicators of knowledge elaboration.

323
Network structure entropy

Network structure entropy can represent the heterogeneity of the target knowledge map (Wu et al., 2007; Xiao et al., 2008; Hwang et al., 2010). Lower heterogeneity indicates that different nodes have an almost equal importance, while higher heterogeneity means that there are significant differences in the importance of nodes (Tan & Wu, 2004). Meanwhile, the higher heterogeneity of a knowledge map also indicates the more differences of elaborating knowledge. Network structure entropy can be calculated using formula (1) (Tan & Wu 2004):

\[ E(G) = -\sum_{i=1}^{N} I_i \ln I_i \]  

(1)

where \( E(G) \) denotes network structure entropy and \( I_i \) denotes the importance of vertex \( i \), with \( I_i = \frac{d_i}{\sum_{i=1}^{N} d_i} \). Additionally, \( d_i \) and \( N \) denote the degree of vertex \( i \) and the total number of vertices, respectively. The degree of a vertex is the number of edges incident with the vertex.

Degree distribution index

The degree distribution index indicates the relevance of knowledge and the connectivity of the target knowledge map (Barabasi & Albert, 1999; Ifenthaler, 2010). Higher relevance and stronger connectivity represent that the group of students have effectively elaborated their knowledge during the collaborative learning process. The degree distribution index can be calculated using formula (2):

\[ D(G) = e^{-\frac{2K \times \sum_{i=1}^{N} I_i \ln I_i}{N}} \]  

(2)

where \( D(G) \) denotes the degree distribution index and \( I_i \) indicates the importance of node \( i \), with \( I_i = \frac{d_i}{\sum_{i=1}^{N} d_i} \). \( K \) denotes the total edges of the target knowledge map. \( N \) denotes the total number of vertices.

Depth of knowledge map

The depth of the knowledge map can indicate the level of elaborating on target knowledge (Lund et al., 2007; Salminen, Marttunen & Laurinen, 2010). The depth can be calculated using formula (3):

\[ \text{Depth}(G) = \sum_{i=1}^{N} \max_{L(P_j)} \text{depth}(K_i) = \sum_{i=1}^{N} e^{\max(L(P_1), L(P_2), L(P_3), \ldots, L(P_j))} \]  

(3)

where \( \text{Depth}(G) \) denotes the depth of the knowledge map. \( K_i \) denotes knowledge sub-maps. \( P_j \) denotes the simple paths of the knowledge sub-map, and \( L(P_j) \) denotes the path length of the simple paths. A path in a graph is a sequence of edges which connect a sequence of vertices. A path with no repeated vertices is defined as a simple path. The path length is the number of edges in the path. \( N \) denotes the number of knowledge sub-maps.

Breadth of knowledge map

The breadth of the knowledge map indicates the scope of knowledge elaboration (Lund et al., 2007; Salminen, Marttunen & Laurinen, 2010). It represents the breadth of the understanding of the subject matter. We adopt the diameter of the knowledge map to compute its breadth, as shown in formula (4):

\[ \text{Diam}(G) = \max_{u \in V(G)} \max_{v \in V(G)} d(u, v) = \max \{ d(u, v) \mid \forall u, v \subseteq V(G) \} \]  

(4)
where Diam (G) denotes the breadth of the knowledge map and \( d(u, v) \) denotes the distance from vertex \( u \) to \( v \). The diameter of a graph is the maximum of \( d(u, v) \) over all vertices \( u \) and \( v \).

**Weighted path length of the activation spanning tree**

A spanning tree is composed of all the vertices and some (or perhaps all) of the edges of a graph (Hassin & Tamir, 1995). The activation spanning tree is generated through activating knowledge by the group of students. As is shown in Figure 1, arcs with arrows and nodes consist of the activating knowledge map, while the straight lines and nodes consist of the spanning tree of this knowledge map. The weighted path length of the activation spanning tree represents the semantic richness of a knowledge map, namely the amount of semantic information contained in the knowledge map. Higher semantic richness represents more amount of semantic information when the group of students elaborate knowledge. It can be calculated using formula (5):

\[
WPL = \sum_{i=1}^{N} W_i L_i
\]

Where WPL denotes the weighted path length of the activation spanning tree and \( W_i \) denotes the weight of vertex \( i \), which equals its activation quantity. The activation quantity of vertex \( i \) equals \( \frac{F \cdot \log(d + 2)^n r}{\log(n^* (D - d + 2))} \) where \( d \) denotes the number of the activated edges, \( D \) denotes the total number of edges that is incident with the vertex \( i \). \( F \) and \( r \) are adjustable parameters. The quantity of activation can measure the semantic properties of knowledge (Zheng et al., 2012). \( L_i \) denotes the path length of vertex \( i \). \( N \) denotes the total number of vertices.

**Figure 1.** The spanning tree when activating knowledge

**Hypotheses**

We assume that the abovementioned indicators can measure the level of knowledge elaboration. Thus, the following five hypotheses are proposed:

- **H1:** The heterogeneity of the target knowledge map is positively related to group performance and prior knowledge of a group.
- **H2:** The connectivity of the target knowledge map is positively related to group performance and prior knowledge of a group.
- **H3:** The depth of the target knowledge map is positively related to group performance and prior knowledge of a group.
- **H4:** The breadth of the target knowledge map is positively related to group performance and prior knowledge of a group.
- **H5:** The semantic richness of the target knowledge map is positively related to group performance and prior knowledge of a group.
Computer-assisted knowledge map analytic tool

This study adopted the computer-assisted knowledge map tool to analyse and measure knowledge elaboration. This analytical tool developed by the authors is a web-based system built with html5 and PHP to draw knowledge maps and calculate the level of knowledge elaboration. The system architecture of this tool is shown in Figure 2.

In order to measure the knowledge elaboration level in collaborative learning, the following three steps need to be conducted with the aid of the knowledge map analytic tool:

First, drawing an initial knowledge map according to the collaborative learning tasks. Each collaborative learning task has a definite learning objective, and the initial map represents the designers’ understanding of the domain knowledge with this objective. The initial knowledge map consists of nodes and edges, where the nodes represent knowledge and the edges represent the mutual relationships of that knowledge. The knowledge map can be drawn based on the knowledge modelling norm (Yang, 2010). Figure 3 shows a portion of an initial knowledge map, where SM denotes symbols, CN denotes concepts, PF denotes principles and formulas, FM denotes formats, PS denotes processes and steps, CS denotes cognitive strategies, and FC denotes facts and cases.

Second, coding and segmenting information flows generated in collaborative learning processes. The coding format of information is defined as: \(<\text{time}><\text{IPL}_i><\text{cognitive level}><\text{information type}><\text{representation format}><\text{knowledge sub-map}>\). Time refers to the start time of each information flow. The subscript “i” in IPL is used to distinguish different learners. That is to say, IPL represents the first learner’s information processing; IPL2 represents the second learner’s information processing, and so on. The cognitive levels of IPL include discriminating, recalling, understanding, and applying. Information types include description of the objectives, context, knowledge semantics, answers and questions, facts and examples, management instructions, related information, and unrelated information. The values of the representation formats include text (T), sound (S), graph (G), photo (P), table (Tb), video (V), animation (A), object (O), and body language (B). The knowledge sub-map mapped by the information flows represents pieces of knowledge and their interrelationships. The initial knowledge map may be revised according to the information flows generated by each group. Table 2 shows the interactional fragments of a group. Two researchers coded these information flows into information items with the aid of the knowledge map analytic tools, as shown in Figure 2. For example, when the information flow in the second row of the Table 2 was coded, the first step was to identify time, IPL, cognitive level, information type, representation format, and knowledge sub-map...
based on the aforementioned coding format. In terms of this information flow, time was 4:59, IPLi was IPL2, cognitive level was “understanding,” information type was “knowledge semantics,” representation format was “sound,” and knowledge sub-map contained two nodes (i.e., curriculum objectives and expected results) and their relationship (i.e., equals) in the initial knowledge map, as shown in Figure 3. The second step was to input the information one by one into the right part of Figure 4 via the knowledge map analytic tool. The coding result was the second information item in the middle part of Figure 4. The final knowledge map was generated after all of the information flows were coded.

![Knowledge Map](image-url)

*Figure 3. A portion of an initial knowledge map*

<table>
<thead>
<tr>
<th>Time</th>
<th>IPLi</th>
<th>Information flows</th>
</tr>
</thead>
<tbody>
<tr>
<td>4:12</td>
<td>IPL1</td>
<td>“The task was to understand the concept of curriculum objectives. Let’s begin!”</td>
</tr>
<tr>
<td>4:59</td>
<td>IPL2</td>
<td>“I believe that curriculum objectives equal the expected results.”</td>
</tr>
<tr>
<td>5:08</td>
<td>IPL2</td>
<td>“For example, the curriculum objective of a course.”</td>
</tr>
<tr>
<td>5:18</td>
<td>IPL3</td>
<td>“Yes, except the curriculum objective of a course, there was the curriculum objectives of a class.”</td>
</tr>
<tr>
<td>5:28</td>
<td>IPL3</td>
<td>“So curriculum objectives equal the standards and requirements of instruction.”</td>
</tr>
<tr>
<td>5:40</td>
<td>IPL3</td>
<td>“In addition, I think curriculum objectives also equal the required knowledge and competences.”</td>
</tr>
<tr>
<td>6:10</td>
<td>IPL1</td>
<td>“Would you like to tell me the relationships between curriculum objectives and instructional objectives?”</td>
</tr>
</tbody>
</table>

Table 2. Fragments of information flows

Third, generating the final knowledge map and computing the level of knowledge elaboration of this knowledge map. We hypothesized that some attributes of the knowledge map can represent the level of knowledge elaboration. When clicking the button of “compute attributes” in Figure 4, the network structure entropy, degree distribution index, depth, breadth, and weighted path length of the activation spanning tree of the target knowledge map can be automatically calculated via the analytical tool. This study therefore aims to validate whether they can measure the level of knowledge elaboration. The numbers next to the knowledge in Figure 5 are the weighted path lengths which...
were calculated with formula (5). For example, if we selected the vertex of “curriculum objectives” as the root vertex, then the path length of vertex “instructional objectives” was 2 and the activation quantity of this vertex was 3.584. Therefore, the weighted path length of vertex “instructional objectives” was 7.17. The analytical tool can calculate the weighted path lengths of all vertices.

Figure 4. Fragments of coding and segmenting

Figure 5. Final knowledge map with weighted path lengths

Compared with previous coding schemes which focused on speech acts and overlooking knowledge construction (Suthers et al. 2010; Stahl, 2011; Zheng et al., 2012), the proposed analytical approach can represent the relationships
between new knowledge and prior knowledge on the knowledge map. The level of knowledge elaboration can also be automatically calculated via our analytical tool. Moreover, this new approach can provide insight into the semantic richness of knowledge elaboration. Therefore, it is an effective method to measure the level of knowledge elaboration.

**Experiment design**

The present study included two experiments that aimed to explore the knowledge elaboration level of different knowledge types at the group level. The type of knowledge in the first experiment was concepts, while the knowledge type of the second experiment was principles. The collaborative learning objective of the first experiment focused on understanding concepts of curriculum objectives, while the learning objective of the second experiment focused on solving problems by applying the theory of consumer behaviour in microeconomics. The experiment procedures were identical in each experiment. However, the collaborative learning tasks, participants, and measuring tools differed. The following section illustrates the collaborative learning tasks, participants, measuring tools, and procedures in detail.

**Collaborative learning tasks**

The collaborative learning tasks of the first experiment are related to the curriculum objectives, and are described as follows:

- How do you understand the concept of curriculum objectives?
- Please describe the strengths and weaknesses of three representations of curriculum objectives.
- How do you determine curriculum objectives? Please describe the procedure in detail.

The collaborative tasks of the second experiment are related to the theory of consumer behaviour in microeconomics. The details of the tasks are as follows:

- What is your opinion of consumer equilibrium? How can you determine the demand curve?
- Case study: When Xiao Yang works 9 hours a day, he is paid 40 RMB per hour. When his hourly wage rises to 60 RMB, he decides to work only 7 hours a day and to spend more time with his family and at leisure. Please describe Xiao Yang’s response to a wage increase using the theory of consumer behaviour.
- The utility function and budget constraint of one consumer can be represented as \( U = \frac{1}{2}Y \) and \( 3X + 4Y = 100 \), respectively. Another consumer’s utility function and budget constraint equations are \( U = X^{0.4} + 1.5\ln X + \ln Y \) and \( 3X + 4Y = 100 \), respectively. Please solve the following problems:
  1. What are the optimal commodity purchase quantities for the two consumers? Please compute them using two methods.
  2. Does “two indifference curves cannot intersect” contradict the conclusion?
  3. If the equation of the inverse demand function of a consumer is \( p = a - b*q \) (\( a, b > 0 \)). When the government imposes a consumption tax, the price paid by this consumer will increase from \( p \) to \( p + pt \), where \( t \) denotes the tax rate. Please prove the following proposition using two methods: The loss of consumer surplus is always more than the revenue of the government from taxation.

**Participants**

The first experiment recruited 91 undergraduates majoring in curriculum theory, while the second experiment recruited 94 undergraduates majoring in economics. In each experiment, the participants were randomly divided into thirty groups of three or four. Thus, a total of 60 groups participated in the two experiments.

**Measuring tools**

In this study, a pre-test and a post-test were conducted to measure the learning performance of individual groups in terms of “curriculum objectives” and “the theory of consumer behaviour in microeconomics”. The test items were
developed by one of the researchers and the two experts in the field of curriculum theory and microeconomics. The purpose of the pre-test was to evaluate the students’ prior knowledge at the group level; that is, the prior knowledge of a group was equal to the average of the group members’ pre-test scores. On the other hand, the post-test aimed to evaluate group performance in the collaborative learning. The test items of the pre-test were the same as those of the post-test.

The group performance was measured based on the group members’ pre-test and post-test scores and was calculated according to formula (6), which represents knowledge gains during collaborative learning based on the measure proposed by Zheng et al. (2012).

\[
X = \frac{P \times (\sum_{i=1}^{N} X_{\text{posttest}}^i - \sum_{i=1}^{N} X_{\text{pretest}}^i)}{N \times \sqrt{CV}}
\]

where \(X_{\text{posttest}}\) and \(X_{\text{pretest}}\) denote the pre-test and post-test total scores of a group, respectively. \(P\) denotes the difficulty coefficient and \(P = 1 - L/W\). \(L\) denotes the mean of loss score of post-test and \(W\) denotes the perfect score. \(CV\) denotes the variation coefficient and \(CV = \frac{S}{\bar{X}} \times 100\%\). \(S\) denotes the standard deviation of difference between the post-test and pre-test in each test item. \(\bar{X}\) denotes the mean of difference between the post-test and pre-test. \(N\) denotes the size of the group. For example, if the \(X_{\text{posttest}}\) and \(X_{\text{pretest}}\) in a group of three were 160 and 60.5, respectively. \(W\) equalled 100 and \(L\) equalled 47, then \(P\) equalled 0.53. \(S\) equalled 4.7 (The differences between the post-test and pre-test in each test item was 19.5, 12.6, 10, 21, and 13) and \(\bar{X}\) equalled 33.17, then \(CV\) equalled 0.14. Therefore \(X\), namely group performance equalled 47.51.

Procedure

The experiment employed a pre-test/post-test research design. The experimental procedures were as follows:

Firstly, the researchers designed the collaborative task and pre-test and post-test items according to target domain knowledge. The target domain knowledge was selected according to collaborative learning objectives. For the first experiment, the target domain knowledge included the concepts, together with categories of curriculum objectives, and methods for determining the curriculum objectives. That is, the target domain knowledge in the first experiment mainly consisted of concepts, while the second experiment was conducted to replicate the results of the first experiment and extended in two ways. First, the results were generalized to collaborative learning tasks that represented different knowledge types. Second, the data from more samples were collected to determine the robustness of knowledge elaboration indicators. Therefore, the target knowledge in the second experiment was extended to include cardinal utility theory, ordinal utility theory, consumer demand theory, consumer equilibrium theory, and consumers’ surplus theory. That is, the target domain knowledge in the second experiment mainly consisted of principles.

Secondly, participants were recruited by posters on the campus. A total of 185 undergraduates voluntarily participated in these experiments. Before collaborative learning, all participants received the same instructions about the purpose and procedures of the experiment.

Thirdly, participants collaborated face-to-face for approximately two hours to complete the tasks in different labs. For each group, group members elected one member to be a group leader. Then the group leader allocated different roles to members based on each member’s abilities or willingness, including information searchers, recorders, analysers, and so on. They could ask for help from the Internet via computers or mobile phones. The final product of each collaborative learning task was a written text. The tasks were the same for the thirty groups in each experiment, as were the pre-test and post-test items. To ensure no interference during the collaborative learning process, we videotaped the entire collaborative learning process for all sixty groups for analysis; moreover, the post-test was performed immediately after the learning activity.
During the collaborative learning processes, we carried out video recording to document each group’s information flows and elaboration processes. In addition, two coders independently segmented information flows generated in collaborative learning processes according to the segmentation rules via our analytical tool. The information flow is segmented if the contributor of information changes. When the cognitive level changes from “discriminating, recalling, and understanding” to “applying”, the information flow is segmented. If the information types change from one type to another, the information flow is also segmented. If the knowledge sub-map changes, it is also necessary to segment the information. However, the change of representation formats does not influence the segmentations because one information flow could have many kinds of representation formats such as sounds, texts or graphs. The attributes of the target knowledge map, such as the network structure entropy, degree distribution index, depth, breadth, and weighted path length of the activation spanning tree were computed by our analytical tool.

To guarantee the objectivity of the information flow coding and test item assessment, two trained raters blind to the experimental conditions coded the information flows of the sixty groups and assessed the test papers of the 185 participants. The percent agreement index was used to compute the inter-rater reliability. The reliability coefficient for coding information flows ranged from 0.900 to 0.974. All inter-rater reliability coefficients for assessing test items were above 0.9. The two raters discussed and resolved all discrepancies. These values of inter-rater reliability were regarded as an indication of excellent agreement. The retest reliability of the two coders for 10% of the data reached 0.93.

In order to test the five hypotheses, descriptive statistical analysis, correlation analysis and regression analysis were conducted for group performance, prior knowledge of a group, and five indicators of knowledge elaboration. Table 3 shows the mean and standard deviation of group performance, prior knowledge, and various indicators. Table 4 shows the regression analysis results for predicting group performance by knowledge elaboration indicators. Both Table 3 and Table 4 show the results of the first experiment.

### Table 3. Descriptive statistics for group performance, prior knowledge and predictors of knowledge elaboration

<table>
<thead>
<tr>
<th>Items</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group performance</td>
<td>19.232</td>
<td>9.748</td>
</tr>
<tr>
<td>Prior knowledge</td>
<td>26.183</td>
<td>8.778</td>
</tr>
<tr>
<td>Network structure entropy</td>
<td>3.723</td>
<td>.245</td>
</tr>
<tr>
<td>Degree distribution index</td>
<td>6.249</td>
<td>.709</td>
</tr>
<tr>
<td>Depth</td>
<td>75.869</td>
<td>34.345</td>
</tr>
<tr>
<td>Breadth</td>
<td>7.270</td>
<td>1.112</td>
</tr>
<tr>
<td>Weighted path length of activation spanning tree</td>
<td>619.66</td>
<td>328.76</td>
</tr>
</tbody>
</table>

### Table 4. Regression analysis for knowledge elaboration indicators predicting group performance

<table>
<thead>
<tr>
<th>Indicators</th>
<th>Adjusted $R^2$</th>
<th>Beta</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network structure entropy</td>
<td>.326</td>
<td>.591</td>
<td>3.879</td>
<td>.001</td>
</tr>
<tr>
<td>Degree distribution index</td>
<td>.259</td>
<td>.533</td>
<td>3.337</td>
<td>.002</td>
</tr>
<tr>
<td>Depth</td>
<td>.132</td>
<td>.402</td>
<td>2.325</td>
<td>.028</td>
</tr>
<tr>
<td>Weighted path length of activation spanning tree</td>
<td>.345</td>
<td>.606</td>
<td>4.036</td>
<td>.000</td>
</tr>
</tbody>
</table>

H1 assumed that the network structure entropy of the target knowledge map was positively related to group performance and prior knowledge. The result indicated that network structure entropy was significantly positively correlated with group performance ($r = .591, p = .001$), and could explain 32.6% of the total variance. This finding indicated that the network structure entropy of the target knowledge map could predict group performance.
significantly well. The main reason for this good performance is the ability of network structure entropy to measure the heterogeneity of the knowledge map. Meanwhile, the network structure entropy of the target knowledge map was positively related to prior knowledge ($r = .714, p = .000$).

H2 assumed that the degree distribution index of the target knowledge map was positively related to group performance and prior knowledge. The result showed that the degree distribution index was significantly positively correlated with group performance ($r = .533, p = .002$). The degree distribution index could explain 25.9% of the total variance, indicating that the degree distribution index of the target knowledge map could predict group performance significantly well. This good performance is attributed to the ability of the degree distribution index to indicate the degree of association among target knowledge and the connectivity of knowledge map. In addition, the degree distribution index of the target knowledge map was also positively related to prior knowledge ($r = .751, p = .000$).

H3 assumed that the depth of the target knowledge map was positively related to group performance and prior knowledge. The result indicated that the depth of the target knowledge map was significantly positively correlated with group performance ($r = .402, p = .028$). The depth of the target knowledge map could explain 13.2% of the total variance, indicating that it could predict group performance significantly well, which is attributed to its ability to represent the elaboration level of knowledge construction. Moreover, the depth of the target knowledge map was positively related to prior knowledge ($r = .449, p = .013$).

H4 assumed that the breadth of the target knowledge map was positively related to group performance and prior knowledge. However, the result indicated that it was not significantly correlated with group performance ($r = .281, p = .133$), which is attributed to its inability to represent a deeper understanding of the subject matter. Furthermore, the breadth of the target knowledge map was not related to prior knowledge ($r = .242, p = .198$).

H5 assumed that the weighted path length of the activation spanning tree of the target knowledge map was positively related to group performance and prior knowledge. The result indicated that the weighted path length of the activation spanning tree of the target knowledge map was significantly positively correlated with group performance ($r = .606, p = .000$), and it could explain 34.5% of the total variance. This finding indicated that it can significantly predict group performance, which is attributed to the ability of the weighted path length of the activation spanning tree to measure the semantic richness and deep structure of the knowledge map. It was also positively related to prior knowledge ($r = .676, p = .000$).

Both Table 5 and Table 6 are the results of the second experiment, which also revealed that the network structure entropy ($r = .483, p = .007$), degree distribution index ($r = .482, p = .007$), depth ($r = .399, p = .029$), and weighted path length of the activation spanning tree ($r = .448, p = .013$) of the target knowledge map were positively related to prior knowledge. In addition, we examined the relationships between these indicators and group performance. Table 5 shows the results for different indicators predicting group performance. The results indicated that H1, H2, H3, and H5 were supported by the data. However, H4 was not supported, indicating that the breadth of the target knowledge map ($M = 11.130, SD = 1.040$) was not significantly positively correlated with group performance ($r = .173, p = .360$) or with prior knowledge ($r = .351, p = .057$). This finding also indicated that the network structure entropy, degree distribution index, depth, and weighted path length of the activation spanning tree of the target knowledge map can significantly predict group performance. It was thus obvious that the conclusions of the second experiment were the same as those of the first experiment. Therefore, we believe that these four indicators are stable and have strong predictive power.

### Table 5. Descriptive statistics for group performance, prior knowledge, and predictors of knowledge elaboration

<table>
<thead>
<tr>
<th>Items</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group performance</td>
<td>16.128</td>
<td>10.885</td>
</tr>
<tr>
<td>Prior knowledge</td>
<td>23.688</td>
<td>7.873</td>
</tr>
<tr>
<td>Network structure entropy</td>
<td>4.003</td>
<td>.176</td>
</tr>
<tr>
<td>Degree distribution index</td>
<td>6.109</td>
<td>.622</td>
</tr>
<tr>
<td>Depth</td>
<td>100.080</td>
<td>46.403</td>
</tr>
<tr>
<td>Breadth</td>
<td>11.130</td>
<td>1.040</td>
</tr>
<tr>
<td>Weighted path length of activation spanning tree</td>
<td>1204.554</td>
<td>527.308</td>
</tr>
</tbody>
</table>

332
The topological properties of knowledge maps were adopted and analyzed using graph theoretical approaches (Hwang, Wu, & Ke, 2011). The network structure entropy, degree distribution index, and depth represented the topological characteristics of targeting knowledge maps. When the weighted path length of the activation spanning tree reflected the semantic richness, it has stronger predictive power than other indicators. Furthermore, knowledge maps display knowledge elaboration processes by representing sequences, branches, and pathways that are involved in collaborative learning.

To sum up, on the one hand, the four indicators, namely the network structure entropy, degree distribution index, depth, and weighted path length of the activation spanning tree can measure the knowledge elaboration level. On the other hand, the knowledge elaboration level can significantly predict group performance. Hence, compared with pre- and post-test, the knowledge elaboration level can be considered a new criterion to measure group performance.

### Discussion and conclusions

In this study, the topological properties of knowledge maps were adopted and analyzed using graph theoretical approaches (Hwang, Wu, & Ke, 2011). The network structure entropy, degree distribution index, and depth represented the topological characteristics of targeting knowledge maps. When the weighted path length of the activation spanning tree reflected the semantic richness, it has stronger predictive power than other indicators. Furthermore, knowledge maps display knowledge elaboration processes by representing sequences, branches, and pathways that are involved in collaborative learning.

It is found that knowledge elaboration indicators can be applied to two different knowledge types. The results of the two experiments confirm that the network structure entropy, degree distribution index, depth, and weighted path length of the activation spanning tree of the target knowledge map are the general indices for measuring the level of knowledge elaboration. However, the breadth of the target knowledge map is not an indicator of knowledge elaboration. Therefore, network structure entropy, degree distribution index, depth, and weighted path length of the activation spanning tree will be effective indicators to measure the knowledge elaboration level.

In line with what has been reported by several previous studies (Stark et al., 2002; Noroozi et al., 2012; Stegmann et al., 2012), this study indicated that the knowledge elaboration values were significantly related to group performance. Therefore, we can predict the group performance of collaborative learning through the lens of knowledge elaboration. Elaborating knowledge entails integrating and refining new information by organising and connecting with prior knowledge. The present study also showed that knowledge elaboration was positively related to prior knowledge. This finding was also congruent with Wetzel (2011) and Van Blankenstein et al. (2013), who found that elaboration was helpful for students with prior knowledge. It was also consistent with Gurlitt & Renkl (2010), who found that elaboration is a key process for prior knowledge activation. In fact, prior knowledge sets the stage for new learning and promotes the processing of new information related to the prior knowledge. The results of this study are also consistent with the definition of knowledge elaboration, which emphasises linking and integrating new knowledge with prior knowledge. In sum, the level of knowledge elaboration was found to be significantly related to group performance and prior knowledge of a group. Thus, researchers or teachers can know about group performance and prior knowledge by measuring knowledge elaboration without a pre-test or post-test.

Methodologically, the knowledge map analytic approach is based on topology characteristics and semantic relationships of knowledge maps, which can provide insights into the processes of knowledge elaboration. We believe that it is knowledge that can be replicated in different conditions, because knowledge is relatively stable but learners are ever-changing. Therefore, the samples in this study are not the participants but the target knowledge map. We attempted to analyse the relationships between the characteristics of the target knowledge map and the elaboration level. This new approach focuses on analysing different attributes of the same knowledge map. This study also contributes to the advancement of the collaborative learning field using the knowledge map-based analysis method.

This study has some implications for educators and practitioners in collaborative learning. Because knowledge elaboration is helpful for meaningful learning and knowledge gains, teachers should provide some external instructional guidance to facilitate knowledge elaboration. Van Boxtel (2004) postulated that elaboration occurred
when students used examples, analogies, and experiences to create new relationships. We believe that the design of collaborative learning tasks should promote relating prior knowledge to new information. Additionally, before collaborative learning, some relevant prior knowledge should be provided to students. During collaborative learning, teachers or facilitators can ask questions to retrieve relevant prior knowledge to promote knowledge elaboration. Teachers should also encourage group members to give more examples for concept learning. After collaborative learning, summarising what was discussed can also stimulate elaboration because students can integrate new knowledge into prior knowledge. Future research is encouraged to explore which kind of knowledge elaboration strategy will be more effective in different collaborative learning scenarios.

Of course, the present study also has some limitations. One limitation is that the learning activity was conducted in a face-to-face collaborative learning context. It would be interesting to investigate whether the result can be generalised to synchronous or asynchronous collaborative learning scenarios. Another limitation is that this study only examined the knowledge elaboration level of two kinds of knowledge type. It is worth exploring the elaboration level of other knowledge types in future studies. Finally, this study mainly focused on four indicators for measuring the level of knowledge elaboration. It is possible that other attributes of knowledge map can be considered to more accurately measure the level of knowledge elaboration in the future.

In conclusion, this study indicates that the network structure entropy, degree distribution index, depth, and weighted path length of the activation spanning tree of the target knowledge map are indicators of knowledge elaboration. It should be noted that the breadth of the target knowledge map is not an indicator since it cannot measure the knowledge elaboration level. Moreover, knowledge elaboration is positively related to both group acquisition of knowledge and prior knowledge of a group. The main contribution of this study resides in the indicators of measuring knowledge elaboration. In addition, an innovative analytic method is developed for analysing knowledge elaboration in collaborative learning. Analysis of the knowledge map can deepen our understanding of the nature of knowledge elaboration processes in collaborative learning. Measuring knowledge elaboration can also provide an insight into prior knowledge and group performance.

Acknowledgements

This study was supported by the Twelfth Five-Year Plan of the National Educational Science in China (ECA110330).

References


Does Gender Influence Emotions Resulting from Positive Applause Feedback in Self-Assessment Testing? Evidence from Neuroscience

Chia-Ju Liu¹, Chin-Fei Huang¹, Ming-Chi Liu², Yu-Cheng Chien², Chia-Hung Lai² and Yueh-Min Huang²*

¹Graduate Institute of Science Education, National Kaohsiung Normal University, Taiwan // ²Department of Engineering Science, National Cheng Kung University, Taiwan // chiжу1105@gmail.com // chinf1027@yahoo.com.tw // liumingchi@gmail.com // yucheng.chien7551@gmail.com // chlain.ncku@gmail.com // huang@mail.ncku.edu.tw

*Corresponding author

(Submitted February 17, 2014; Revised May 18, 2014; Accepted June 25, 2014)

ABSTRACT

Computerized self-assessment testing can help learners reflect on learning content and can also promote their motivation toward learning. However, a positive affective state is the key to achieving these learning goals. This study aims to examine learning gains and emotional reactions resulting from receiving emotional feedback in the form of applause during computerized self-assessment testing of university students. The participants were asked to solve mathematics problems in a computer-assisted self-assessment system with or without pre-recorded applause as emotional feedback while EEG measurements were taken. Using psychological evidence from neuroscience technology, we tested the hypothesis that the part of the brain that generates feelings of reward is more active in male students than in female students during computer-assisted self-assessment testing. The results of this study provide support for the belief that it is useful to reduce negative emotional states in students by using emotional reactions such as applause during computer-assisted self-assessment testing, especially in the case of male students. It is suggested that instructors may wish to create such a positive emotional self-assessment learning environment to encourage students to learn by themselves more efficiently.

Keywords

Evaluation methodologies, Gender studies, Interactive learning environments, Teaching/learning strategies

Introduction

As one of the indispensable instruments in the educational domain, assessment is often carried out by teachers to measure students’ learning gains. From the perspective of students, however, assessment may likely turn into stress and anxiety if they fail to achieve expected grades (Caraway, Tucker, Reinke, & Hall, 2003; Huang, Huang, & Wu, 2014). To this end, self-assessment test systems have been typically considered as an effective instructional strategy for training students not only to evaluate their own learning progress, but also to help them relieve anxiety and other negative emotional states while facing real tests (Moridis & Economides, 2012). Indeed, anxiety relief is a positive factor for real test achievement, which may be attributable to training in self-assessment (Grete & Green, 2000; Snooks, 2004). Moreover, through self-assessment, students are able to review the course content being taught, retrieve what they have learned and even promote reflection (Kulik, Kulik, & Bangert, 1984).

Although self-assessment is beneficial for students with regard to enhancing what they have learned through review and reflection, it often creates a stressful experience similar to that carried out by paper-based tests (Cassady, 2004; Ricketts & Wilks, 2002). The stress is presumably linked to the fact that the majority of real tests most students face are still paper-based. Studies have indicated that computer-based assessments, hereafter CBAs, with a user-friendly interface are a sound alternative for self-assessment purposes (Kaklauskas et al., 2010; Ricketts & Wilks, 2002). Thanks to information technology, today’s CBAs can not only provide instant feedback but can also adaptively generate questions for students, reducing the time and cost of such assessments (Terzis & Economides, 2011). As a consequence, CBAs become an ideal platform by which learners can engage in self-assessment (Moridis & Economides, 2012; Nicol & Macfarlane-Dick, 2006). Due to the adaptivity of this method, computerized self-assessment tests have been widely adopted.

Research has supported the argument that computerized self-assessment testing can provide efficient feedback to help improve student reflection and learning motivation, and it can promote higher levels of learning achievement (Black & William, 1998; Wilson, Boyd, Chen, & Jamal, 2011). However, other studies have indicated that students...
Affect is the basis of human experience. There are two types of affect, positive and negative (Economides & Moridis, 2008), where positive affect (e.g., acceptance, joy, confidence, etc.) has a positive impact on learning, memory and thinking, resulting in learning within an optimal range of human functioning. On the contrary, negative affect (e.g., anxiety, anger, fear, sadness, etc.) has a negative impact on motivation and leads to inattention and even distraction. Moridis and Economides (2012) and Huang and Liang (in press) indicated that negative affect adversely impedes learners. Negative affect occurs most when the students' final learning outcomes do not meet their expectations, such as when they answer a question wrongly or fail a test. As affect can significantly influence learning, understanding learners' affect throughout the learning process is crucial for understanding their learning motivation. However, it is impossible to put learners' affect into perspective and to give them immediate feedback when it comes to paper-based assessments (Sung, Chang, Chiou, & Hou, 2005).

Issues of immediate affective feedback for computer-assisted self-assessment have been increasingly discussed in recent studies (Liu, Papathanasiou, & Hao, 2001; Moridis & Economides, 2012). These studies all have suggested that immediate affective feedback for computer-assisted self-assessment can help promote student self-confidence and improve performance. Furthermore, they all confirmed that positive affective feedback for computer-assisted self-assessment can decrease students’ anxiety and can improve their levels of learning achievement.

Differences in terms of the behaviors or reactions of individuals using various computer applications, including computer-assisted learning, may be due to gender. One study indicated that the participation of females in ICT professional careers is low and even still falling in most western countries (Meelissen & Drent, 2008). Thus, an experiment was conducted to investigate those factors related to elementary schools or teachers in regard to their influence on female students’ attitudes toward computers. The results revealed that female children's attitudes toward computers were less positive than that of males. Will this be the same when it comes to the issue of affect? It is a very interesting issue and worthy of investigation.

Moridis and Economides (2012) adopted applause as an achievement-based reward during a computer-assisted self-assessment test. They found that males who were not receiving applause during the test had a significantly higher state of anxiety than females who were not receiving applause either. The state of anxiety assessed in this study using the scores of the State-Trait Anxiety Inventory (STAI) was based on the students’ self-perceptions after the test, hardly realizing how continuously the state of emotion may have been changing during the test. In fact, we believe affect is a state and should therefore be observed instantaneously. Therefore, both the STAI and objective scientific evidence are required to explore the issue of affect.

In this study, we tackle gender differences in affect and explore why feelings of reward influence students’ test performance using neuroscience technology, namely Electroencephalography (EEG). The research questions are as follows:

- What are the possible influences of applause on test performance?
- What are the differences in anxiety, if any, between different genders when stimulated by applause?
- To what extent can applause affect anxiety?

In this study, a neuroscience technology, an EEG, was used to measure the effect of applause on anxiety. Huang and Liu (2012) reported that EEG reflects the corresponding patterns of cognition or emotion in participants. Thus, EEG can precisely measure participant anxiety.

It is intended that the results of this study will provide psychological evidence by which to interpret the underlying reasons for the findings as well as useful suggestions for the design of more comprehensive self-assessment test systems. Furthermore, the results could serve as a reference to guide teaching and learning strategies regarding emotional reactions during computer-assisted self-assessment testing.

The purpose of this study is to provide psychological evidence using neuroscience technology to test the hypothesis that the part of the brain that generates feelings of reward is more active in males than in females during computer-
assisted self-assessment testing. The findings could reflect the effect of emotional reactions on learning gains by receiving applause during a computer-assisted self-assessment test.

Methodology

Test design

The Hot Potatoes System was adopted in this study to design the computer-self-assessment tests. The design process is shown in Figure 1. In the first step, 15 single-choice questions were chosen to compose the tests for both the experimental and controlled tasks. Second, the mathematics questions administered in the tasks were a series of two-digit additions. The simple math problems were aimed at eliminating the influence of participants’ prior knowledge on their test achievement. Moreover, participant reactions evoked by complex math problems might potentially involve various levels of cognitive processing, and therefore, the main effect of applause on the test achievement investigated in this study would be difficult to explain from the results of such complex math problems. This task design (two-digit additions questions for undergraduates) was also employed in studies conducted by Cates and Rhymer (2003) and Sheffield and Hunt (2006). The choice of 15 single choice tests in both the experimental and controlled task was a result of concern about the participants' attention processing load. As suggested by general EEG methodology, people can pay the best sustained attention for 5 minutes. Third, the test carried out in the experimental task used the applause sound for right answers to questions, whereas the test carried out in the controlled task did not. Fourth, the web page-based tests for the experiment were generated by the Hot Potatoes system.

![Figure 1. The design for the test process using the Hot Potatoes System](image1)

The flow for the experimental task is shown in Figure 2. In the experimental task, a pre-recorded sound of applause was embedded into the question feedback. If a student answered the question correctly, the system instantly reacted with the applause sound, accompanied by the word “correct” shown on the computer screen (Figure 2a). If the student got the wrong answer, no applause sound was emitted, and the word “wrong” was shown on the screen (Figure 2b). In the controlled task, there was no applause sound played for the correct answer; however, the word “correct” still appeared on the screen (Figure 3a). If the student got the wrong answer, the word “wrong” was shown (Figure 3b), as in the experimental task. Finally, the computer-assisted self-assessment tests for the experimental and controlled tasks were exported for the achievement tests. It should be noted that the tests were developed in Chinese.

![Figure 2. The design of the computer-assisted self-assessment test for the experimental task](image2)

(a) The sound of applause is played with the word “correct” displayed on the screen when the students answer correctly (b) The word “wrong” is shown on the screen when the students get the wrong answer
Participants

This study was conducted at an urban university in Taiwan. Thirty participants ($n = 30$, 15 males, 15 females; mean age $\pm SD = 19.2 \pm 2.0$ years) participated in this experiment and were divided into two groups: one male group and one female group. All participants were asked to complete both the experimental and controlled task tests throughout the whole process to collect the psychological data. All participants were confirmed to be mentally healthy without a history of neurological or psychiatric disorders, and all gave voluntary consent to participate in the neuroscience experiments. This study conformed to The Code of Ethics of the World Medical Association (Declaration of Helsinki) and was approved by the ethics committee of National Kaohsiung Normal University.

Using an EEG (Electroencephalography) experiment to provide neuroscience evidence

Many cognitive processes are difficult to explain verbally (Bargh & Ferguson, 2000; Huang, Liu, Chen, Kinshuk, & Wen, 2014; Liu, Huang, & Wen, 2013); hence, some studies have suggested that the study of cognitive processes must combine neurophysiological methods with questionnaires (Wang & Chiew, 2010). Past neuroscience studies have found that humans show similar trends while responding to the same task in elicited EEG data, such as that derived during recognition or problem solving tasks (Cicconetti et al., 2007; Juckel et al., 2008; Lai, Lin, Liou, & Liu, 2010) and have also suggested that the EEG research method is suitable for exploring research issues related to emotions and cognitive processes.

In this study, the EEG supplied neuroscience data indicative of the effect of gender differences on emotional reactions during a computer-assisted self-assessment test, and to the best of our knowledge, this is the first study to provide concrete evidence regarding this issue. The participants were required to complete both the experimental and controlled task tests while neuroscience technology was used to collect the psychological data throughout the entire process. EEG is a procedure that measures the electrical activity of the brain through the skull and scalp (see Figure 4) (Rugg & Coles, 1995). For collecting the electrical activity of the brain, the participants were required to put on the electrode cap before doing the test. Figure 4 indicates the relative position of the different areas of the human brain on the cap. The standard for the relative position of the electrode cap in this study was adopted from the international 10-20 system. The character “F” means frontal lobe; the character “P” means parietal lobe; the character “C” means central lobe; the character “T” means temporal lobe, and the character “O” means occipital lobe. When participants experience emotional reflections, the corresponding electrical activities in the brain are induced (Huang & Liu, 2012).
The EEG was amplified (band pass, 0.01-40 Hz) using SynAmps/SCAN 4.4 hardware and software (NeuroScan, Inc., Herndon, VA) and a commercial electro-cap (Electro-Cap International, Eaton, OH) with electrodes at 32 scalp locations based on the 10-20 International system (see Figure 5).

The noise signals were filtered out automatically. The electrode impedance was kept below 5 kΩ. The averaging epoch was 1,024 ms, including 200 ms of pre-stimulus baseline. EEG channels were continuously digitized at a rate of 1,000 Hz using a SynAmp™ amplifier. The signal was analogy filtered (0.1-200 Hz), A/D converted with a sampling rate of 1,000 Hz, 14 bit precision and digitally filtered in the range of 0.1-30 Hz. All of the participants’ EEG signals were recorded while they were completing the experimental tasks (Figure 6).
The procedure and data collection

All of the students were advised that they were to take a computer-assisted mathematics test in this study (see Figure 7). They were also told that they could participate in a computer-assisted self-assessment test as practice before the test. All students in this study took the computer-assisted self-assessment test twice in a different order. One half of them took the controlled test first, which was then followed by the experimental test. The other half took the tests in the reverse order. The two tests were separated by an interval of one hour. The duration of each test was approximately 23 minutes. Students were asked to answer each question within four seconds. The state of anxiety questionnaire was distributed before and after the control tests. Also, the state of anxiety was written down before and after the experimental tests. The extracted data were then analyzed using a repeated-measures ANOVA analysis.

![Timeline for both the controlled and experimental tasks](image)

*Figure 7.* Timeline for both the controlled and experimental tasks

In the process of completing the tests, all participants had to wear a 32-channel electrode cap in order to collect the EEG data (see Figure 8a). The process for wearing the electrode cap includes the following steps: First, the cap is placed on the participant's head, and the electrode holders are filled with electrode gel using a syringe. Second, the electrodes are plugged into the cap one by one. The procedure for putting on the electrode cap takes about 30 minutes. At the beginning of the EEG experiment, all students were asked to sit down and relax. The EEG for their rest state was collected as an individual brain wave baseline. Moreover, in order to avoid the participant's emotional state being influenced by the previous test, they were asked to relax and do nothing for 60 minutes between the two tests.

![Wearing the electrode cap and answering the questions](image)

*Figure 8.* The EEG instrument and the experiment in a EEG laboratory
All of the participants’ EEG signals were collected while they were participating in the experiments (see Figure 8b). Frequency analysis was performed in the delta (1–4 Hz), theta (4–7 Hz), alpha 1 (8–10 Hz), alpha 2 (11–14 Hz), beta (15–30 Hz) and gamma (> 30 Hz) frequency bands. The evidence of brain activity from these frequencies and power values could help to verify the hypothesis that the part of the brain that generates feelings of reward is more active in males than in females during the computer-assisted self-assessment test. The most important power values in this study are alpha 2 and delta frequencies. Blackhart, Kline, Donohue, LaRowe, and Joiner (2002) indicated that the power values of the alpha 2 frequency reflect the induced participants’ positive emotions. In addition, the delta frequency in the frontal lobe of brain reflects high-level cognitive processing occurring in a participant (Ho et al., 2012).

Measurement scales

In this study, students' anxiety levels were measured by using the State-Trait Anxiety Inventory (STAI) (Spielberger, 2005) which was the same instrument used in the study of Moridis and Economides (2012). The self-report inventory included 20 items intended to assess the participants’ state of anxiety. Responses to the items ranged from 1 to 4, as follows: (1) not at all; (2) somewhat; (3) moderately so, and (4) very much, according to the students’ feelings. Scores range from a minimum of 20 (lowest anxiety) to a maximum of 80 (highest anxiety). The validity of the contents of the Chinese version was assessed by three professional psychologists, and Cronbach’s alpha was 0.91 for the state subscale.

Two mathematics tests on the addition of two-digit numbers were administered in this study, consisting of 15 single choice type questions with a perfect score of fifteen. The correlation of these two tests was 0.96 (The data was analyzed using a pilot study), and the Cronbach’s alpha values were 0.84 and 0.87 for the two tests. The high correlation indicates that these two tests are highly consistent with each other.

Results and discussion

Computer-assisted self-assessment tests with and without applause were administered in this study, and the statistical data from the state of anxiety questionnaire were collected before and after each test.

The students' test performance and anxiety in the controlled test

The first to be administered was the controlled test without applause as emotional feedback. The results show that the male group scored 12.4 ± 1.5 points compared with 11.3 ± 1.8 points for the female group in the controlled task mathematics test. There was no significant difference between the two groups’ performance on the test. Furthermore, Table 1 shows that there was less difference in the state of anxiety of the female group before and after the test without applause. However, for the male group, the state of anxiety after the controlled test was higher than it was before the test.

| Table 1. Anxiety pre- and posttest mean scores, standard deviations results by gender in the controlled task |
|---------------------------------------------|-------------|-------------|-------------|-------------|
|                                           | Pretest     |             | Posttest    |             |
|                                           | M           | SD          | M           | SD          |
| State of anxiety                          |             |             |             |             |
| Sex                                        |             |             |             |             |
| Female (n = 15)                            | 20.27       | 6.89        | 19.93       | 6.97        |
| Male (n = 15)                              | 16.13       | 3.70        | 18.07       | 3.13        |

Then, repeated-measures ANOVA was used to assess the differences in the scores of the state of anxiety of the two gender groups before and after the controlled computer-assisted self-assessment test (see Table 2). The results show that there was no significant main effect for anxiety scores across gender groups: $F(1,28) = 2.39, p = .13$. This finding is in line with the results of the research by Moridis and Economides (2012), who suggested that the main effect of gender on state of anxiety is not significant. There was also no significant main effect across pre- and posttest time points: $F(1, 28) = 2.82, p = .10$. However, there was a significant main effect across the interaction between gender group and time point: $F(1,28) = 5.67, p = .02$. 343
Table 2. Analysis of variance results for both genders and pre-posttest for anxiety in the controlled task

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>1</td>
<td>135.00</td>
<td>2.39</td>
<td>.08</td>
</tr>
<tr>
<td>Time (pre/posttest)</td>
<td>1</td>
<td>9.60</td>
<td>2.82</td>
<td>.09</td>
</tr>
<tr>
<td>Gender x Time (pre/posttest)</td>
<td>1</td>
<td>19.27</td>
<td>5.67*</td>
<td>.17</td>
</tr>
</tbody>
</table>

*p < .05.

Figure 9 shows that the females had a higher state of anxiety before the test than the males. However, after completing the test without the applause affective feedback, the females showed a state of less anxiety, while the males still showed a significant increase in anxiety. Colley (2003) pointed out that female students tend to have more negative attitudes related to the use of computers than male students, which may explain why the female students had higher levels of anxiety before the computer-assisted self-assessment test. In addition, for some female students, mathematics represents a critical filter to keep females out of science learning (Matteucci & Mignani, 2009), which may be another reason why the female students had higher anxiety than the males. Finally, compared with male students, female students tend to prefer less competitive or challenging tests (McLean & Anderson, 2009). Therefore, the female students had more negative emotions before taking the computer-assisted self-assessment test with mathematics content than the male students.

![Figure 9. Mean anxiety scores across time points (pre/posttest) by gender in the controlled task](image)

However, as shown in Figure 9, during the controlled experiment without applause as affective feedback, the female students exhibited decreased anxiety, while the male students exhibited increased anxiety after completing the test. Some researchers have suggested that females do relatively worse on tests than males because of having lower confidence (Klein, 2007; Matteucci & Mignani, 2009). According to our findings, we propose that the female students were more anxious than the males before doing the test, but they felt more relaxed after completing the test since the operation of the computer was not difficult, and the participants needed only to click the correct answer by using the mouse. We additionally speculate that perhaps the males had lower anxiety than the females because the male students initially had higher self-confidence in their ability to complete the mathematics test using computers. A possible reason why the male students had higher anxiety after testing might be linked to the feeling that they faced competition from other students (Moridis & Economides, 2012). Therefore, the higher anxiety of male students after taking the test might be due to their expectations for high achievement.

344
The students' test performance and anxiety in the experimental test

The second test to be administered was the experimental test, which provided applause as positive feedback after each correct answer. The results of this test showed that the male group scored $13.9 \pm 2.6$ points compared with $13.5 \pm 2.7$ points for the female group. Although the scores for the experimental task for both the male and female groups were higher than the scores for the controlled task, there was no significant difference between the gender groups.

Moreover, Table 3 shows that there is less difference in the state of anxiety for the female group before and after the computer-assisted self-assessment test with applause. However, for the male group, the state of anxiety after the computer-assisted self-assessment test was found to be dramatically lower than the state before the test.

### Table 3. Anxiety pre and posttest mean scores, standard deviations results by gender in the experimental task

<table>
<thead>
<tr>
<th>State of anxiety</th>
<th>Pretest M</th>
<th>Pretest SD</th>
<th>Posttest M</th>
<th>Posttest SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female ($n = 15$)</td>
<td>20.13</td>
<td>7.37</td>
<td>19.47</td>
<td>7.11</td>
</tr>
<tr>
<td>Male ($n = 15$)</td>
<td>16.20</td>
<td>3.26</td>
<td>8.00</td>
<td>3.05</td>
</tr>
</tbody>
</table>

The repeated-measures ANOVA was conducted to assess the differences in the state of anxiety between the male and female groups before the experimental test. The results show there to be no significant difference in their scores ($t = 1.89, p > .05$).

Moreover, the scores were analyzed using repeated-measures ANOVA to assess the differences in the state of anxiety between the male and female groups before and after the computer-assisted self-assessment test (see Table 4). The results show there to be a significant main effect for anxiety scores across gender group: $F(1,28) = 14.94, p < .01$. The anxiety scores were significantly higher for the female groups than for the male group in the experimental task. There was also a significant main effect across pre- and posttest time points: $F(1, 28) = 104.67, p < .001$. The anxiety scores were significantly higher at the pretest time point than at the posttest time point. Moreover, there was a significant main effect across the interaction between gender group and time point: $F(1,28) = 75.56, p < .001$.

### Table 4. Analysis of variance results by gender and pre-posttest of anxiety in the experimental task

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>$F$</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Between subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>1</td>
<td>889.35</td>
<td>14.94**</td>
<td>.35</td>
</tr>
<tr>
<td><strong>Within subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time (pre/posttest)</td>
<td>1</td>
<td>294.82</td>
<td>104.67***</td>
<td>.79</td>
</tr>
<tr>
<td>Gender x Time (pre/posttest)</td>
<td>1</td>
<td>212.82</td>
<td>75.56***</td>
<td>.73</td>
</tr>
</tbody>
</table>

*p < .01. **p < .001.

Figure 10 shows that the males had a lower state of anxiety after the experiment with applause than the females. This finding is also similar to the results of Moridis and Economides (2012), which indicated that males experienced lower anxiety when they received encouragement. Beyer and Bowden (1997) noted that males tend to be concerned about their capacity to deal with the external environment, while females tend to underestimate themselves. This comment is echoed in the findings of this study, which indicate that male student performance is influenced by the external environment, while female student performance is influenced by their own feelings. Here, the applause could be regarded as an external factor.

The students' EEG data in the experimental test

Our findings obtained by the anxiety questionnaire are mostly in line with those found in the literature. However, we argue that anxiety should be regarded as a continuous state rather than as an outcome of a period or event. Therefore, the data from the questionnaire following the test, which presumably measures and reports the subjects' emotional state, needs to be confirmed. In short, continuous observation is a substantial step to by which complement the
questionnaire. For this purpose, the EEG data for the males and females who performed the experimental computer-assisted self-assessment test with applause were analyzed. A topographical map of the brain is shown in Figure 11. For the male group, the images reveal that the power values of the alpha 2 frequency were more active on the two sides of the frontal lobe than they were for the female group. Blackhart et al. (2002) explained that the power values of the alpha 2 frequency from the two sides of the frontal lobe are often induced by the appearance of positive emotions. In other words, higher power values of the alpha 2 frequency mean more positive emotion. Therefore, the findings indicate that in the experimental computer-assisted self-assessment test with applause, the part of the brain that generates feelings of reward is more active in males than in females during computer-assisted self-assessment testing.

Figure 10. Mean anxiety scores across time points (pre/posttest) by gender group in the experimental task

Figure 11. Topographical map of the Alpha 2 frequency of the brain (The indicators of positive affective state: In each frequency, more areas changed into red colors indicated more active of the brain in this frequency domain)
Huang and Liu (2012) reported that the frontal lobe of the human brain dominates high-level thinking, mental rotation and math calculations. Ho et al. (2012) also indicated that the delta frequency in the frontal lobe of the human brain is related to high-level cognitive processing. For this reason, the delta frequency power values in the frontal lobe (F4 electrode) were further analyzed, as shown in Figure 12. The results show that both the male and female groups had higher delta power values when completing the controlled computer-assisted self-assessment test without applause feedback than they did for the test with applause feedback (see Figure 12). That is to say, both the male and female students needed to exert more power value to complete the computer-assisted self-assessment test without emotional feedback.

![Figure 12. The EEG power values](image)

Figure 12 further illustrates that applause as emotional feedback influenced the males more than the females. The higher power value indicates that the participants needed to recruit more brain resources to complete the task. In other words, the higher power value implies that the participants felt the task was difficult. In Figure 12, it can be seen that the males had the highest delta frequency power value (1–4 Hz) when the test was performed without applause (pink line), and the lowest value when with applause (red line). Although the females also had higher power values when the test was performed without applause (green line) than with applause (blue line), the gap between the power values of the two tests is smaller for the female group than for the male group. These findings may provide psychological evidence to support the argument that emotional feedback influences males more than females in computer-assisted self-assessment testing, especially in terms of encouraging positive emotions. The psychological evidence is in line with the results shown in Table 2, indicating that males had a lower state of anxiety after the experiment with applause than the females. By combining the self-report questionnaire with the psychological measures, the reliability of the self-report is supported by the results of the brain wave data. Moreover, the psychological measures can provide comprehensive observations and deeper insights into the learning process.

**Conclusions**

The purpose of this study was to provide psychological evidence based on neuroscience technology to explain the reasons for gender differences and why feelings of reward influence students’ affective states. To the best of our knowledge, this is the first study to provide concrete evidence regarding this issue.

Although the results of the test performances showed there to be no significant difference between the gender groups in either the controlled or experimental tests, there are three main findings about the affective states under
consideration in this study. Firstly, the females experienced significantly higher anxiety before the computer-assisted self-assessment test without applause than the males, but showed a decreasing state of anxiety compared with the males' significantly increasing state of anxiety after completing the test. The changes in the state of anxiety in the case of the female students might be because they were initially nervous about a computer-based mathematics test, but then found that it was relatively easy. On the other hand, we suggest that the males had lower anxiety because they had higher self-confidence in their ability to complete the test than was the case for the females. However, the male students experienced higher anxiety after completing the test perhaps due to their concern about their capacity with respect to the external environment.

Secondly, the males had significantly lower anxiety than the females after the experimental computer-assisted self-assessment test with applause as affective feedback. This finding confirms the first finding in this study suggesting that male student performance is influenced by the external environment, whereas female student performance is influenced by their own feelings.

Thirdly, the results from the psychological data indicate that the part of the brain that generates feelings of reward is more active in males than in females during computer-assisted self-assessment testing with applause. These findings suggest that positive emotional feedback influences males more than females in such a testing scenario. We therefore suggest that instructors need to design different types of feedback intended to address gender-based differences. Instructors should consider that although positive emotional feedback in computer-assisted self-assessment testing can enhance student learning, it is more beneficial for male students than it is for female students. Hence, instructors need to design other feedback or strategies to promote female student learning achievement.

To conclude, the results of this study suggest that the use of emotional reactions such as applause may improve students’ learning states during computer-assisted self-assessment testing, especially in the case of male students. This study suggests that computerized self-assessment testing can integrate a positive emotional feedback mechanism to encourage students so as to promote their learning states and increase motivation. In addition, this study asserts that neuroscience technology can provide substantial and necessary evidence to confirm the influences of emotional feedback on learning. Nevertheless, there are some limitations in this study. In our experiment, we adopted a mathematics test as the learning content of the task. Many previous studies have mentioned that female students dislike mathematics more than male students. The reaction to the mathematics test should therefore be considered as a variable affecting the gender differences in this study. We thus suggest that researchers explore gender differences in emotional reactions during computer-assisted self-assessment testing for different subjects and tasks.

Acknowledgments

This research is partially supported by the "International Research-Intensive Center of Excellence Program" of NTNU and Ministry of Science and Technology, Taiwan, R.O.C. under Grant no. NSC 103-2911-I-003-301, MOST 103-2511-S-006-007-MY3, and MOST 103-2511-S-006-002-MY3.

References


The absence of external evaluative (3), 207
iterature review.
Liu, M., Papathanasiou, E., & Hao, Y.
10.1007/s11423
by location
Liu, M.
disease.


