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 The articles should be original, unpublished, and not in consideration for publication elsewhere at the time of submission to Educational Technology & Society and three months thereafter.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 Fominichov, K. E. Tsiolkovsky Russian State Tech Univ, Russia; Olga S Fominichova, Studio "Culture, Ecology, and Foreign Languages". Russia; Piet Kommers, University of Twente, The Netherlands; Chul-Hwan Lee, Inchon National University of Education, Korea; Brent Muirhead, University of Phoenix Online, USA; 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-Rifon, University of Vigo, Spain; Gautam Biswas, Vanderbilt University, USA; Rosa Maria Bottino, Consiglio Nazionale delle Ricerche, Italy; Mark Bulen, University of British Columbia, Canada; Tak-Wai Chan, National Central University, Taiwan; Kuo-En Chang, National Taiwan Normal University, Taiwan; Ni Chang, Indiana University South Bend, USA; Yam San Chee, Nanyang Technological University, Singapore; Sherry Chen, Brunel University, UK; Bridget Cooper, University of Sunderland, UK; Darina Dicheva, Winston-Salem State University, USA; Jon Dron, Athabasca University, Canada; Michael Eisenberg, University of Colorado, Boulder, USA; Robert Farrell, IBM Research, USA; Brian Garner, Deakin University, Australia; Tiong Goh, Victoria University of Wellington, New Zealand; Mark D. Gross, Carnegie Mellon University, USA; Roger Hartley, Leeds University, UK; J R Isaac, National Institute of Information Technology, India; Mohamed Jenni, University of Tunis, Tunisia; Mike Joy, University of Warwick, United Kingdom; Athanasis Karoulis, Hellenic Open University, Greece; Paul Kirschner, Open University of the Netherlands, The Netherlands; William Kiemun, Texas A&M University, USA; Rob Koper, Open University of 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; Rory 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; Slavi Stoyanov, Open University, The Netherlands; Timothy Teo, Nanyang Technological University, Singapore; Ching-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.

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@massey.ac.nz (Subject: Submission for Educational Technology & Society journal).
# Table of Contents

## Special issues articles

**Guest Editorial: Game Based Learning for 21st Century Transferable Skills: Challenges and Opportunities**
*Francesco Bellotti, Rosa Maria Bottino, Baltazar Fernández-Manjón and Rob Nadolski*
1–2

**Learning When Serious: Psychophysiological Evaluation of a Technology-Enhanced Learning Game**
*Ben Cowley, Martino Fantato, Charlene Jennett, Martin Ruskov and Niklas Ravaja*
3–16

**Game-Based Assessment of Persistence**
*Kristen E. DiCerbo*
17–28

**Undertaking an Ecological Approach to Advance Game-Based Learning: A Case Study**
*Manta Shah and Aroutis Foster*
29–41

**An Investigation of the Interrelationships between Motivation, Engagement, and Complex Problem Solving in Game-based Learning**
*Deniz Eseryel, Victor Law, Dirk Ifenthaler, Xun Ge and Raymond Miller*
42–53

**Multi-User Virtual Environments Fostering Collaboration in Formal Education**
*Nicoletta Di Blas and Paolo Paolini*
54–69

## Invited article

**E-learning in School Education in the Coming 10 Years for Developing 21st Century Skills: Critical Research Issues and Policy Implications**
*Siu Cheung Kong, Tak-Wai Chan, Patrick Griffin, Ulrich Hoppe, Ronghui Huang, Kinshuk, Chee Kit Looi, Marcelo Milrad, Cathleen Norris, Miguel Nussbaum, Mike Sharples, Wing Mui Winnie So, Elliot Soloway and Shengquan Yu*
70–78

## Case study article

**Changing Pedagogy for Modern Learners – Lessons from an Educator’s Journey of Self-Reflection**
*Terry Taylor*
79–88

## Full length articles

**Collaborative mLearning: A Module for Learning Secondary School Science**
*Dorothy DeWitt, Saedah Siraj and Norlidah Alias*
89–101

**Facilitating English-Language Reading Performance by a Digital Reading Annotation System with Self-Regulated Learning Mechanisms**
*Chih-Ming Chen, Jung-Ying Wang and Yen-Chang Chen*
102–114

**Mathematics Synchronous Peer Tutoring System for Students with Learning Disabilities**
*Mengping Tsuei*
115–127

**A Jigsaw-based Cooperative Learning Approach to Improve Learning Outcomes for Mobile Situated Learning**
*Yueh-Min Huang, Yi-Wen Liao, Shu-Hsien Huang and Hsin-Chin Chen*
128–140

**The Integration of Concept Mapping in a Dynamic Assessment Model for Teaching and Learning Accounting**
*David Tawei Ku, Ju-ling Shih and Shu-Huan Hung*
141–153

**A Novel Agent-Supported Academic Online Examination System**
*Tuğrul TAŞCI, Zekeriya PARLAK, Alpaslan KİBAR, Nevzat TAŞBAŞI and H. İbrahim CEBECİ*
154–168

**A Five-Phase Learning Cycle Approach to Improving the Web-based Problem-Solving Performance of Students**
*Fan-Ray Kuo and Gwo-Jen Hung*
169–184
<table>
<thead>
<tr>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic Factors, TPACK Constructs, and Teachers’ Perceptions of Constructivist-Oriented TPACK</td>
<td>185–196</td>
</tr>
<tr>
<td>Joyce Hwee Ling Koh, Ching Sing Chai and Ching-Chung Tsai</td>
<td></td>
</tr>
<tr>
<td>An Interactive and Collaborative Approach to Teaching Cryptology</td>
<td>197–205</td>
</tr>
<tr>
<td>Sasa Adamovic, Marko Sarac, Miladen Veinovic, Milan Milosavljevic and Aleksandar Jevremovic</td>
<td></td>
</tr>
<tr>
<td>Adaptive vs. Fixed Domain Support in the Context of Scripted Collaborative Learning</td>
<td>206–217</td>
</tr>
<tr>
<td>Anastasios Karakostas and Stavros Demetriadis</td>
<td></td>
</tr>
<tr>
<td>An Investigation of Reading Rate Patterns and Retrieval Outcomes of Elementary School Students with E-books</td>
<td>218–230</td>
</tr>
<tr>
<td>Tsung-Ho Liang and Yueh-Min Huang</td>
<td></td>
</tr>
<tr>
<td>The Effect of Intrapsychology Learning Before and After Interpsychology Activities with a Web-based Sharing Mechanism</td>
<td>231–247</td>
</tr>
<tr>
<td>Guo-Liang Hsu and Wu-Yuin Hwang</td>
<td></td>
</tr>
<tr>
<td>The Effect of Instructional Techniques on Critical Thinking and Critical Thinking Dispositions in Online Discussion</td>
<td>248–258</td>
</tr>
<tr>
<td>Filiz Kalelioğlu and Yasemin Gülbahar</td>
<td></td>
</tr>
<tr>
<td>Adaptive Instruction to Learner Expertise with Bimodal Process-oriented Worked-out Examples</td>
<td>259–271</td>
</tr>
<tr>
<td>Jihyun Si, Dongsik Kim and Chungsoo Na</td>
<td></td>
</tr>
<tr>
<td>A Model-Based Behavior Analysis Approach for Open-Ended Environments</td>
<td>272–282</td>
</tr>
<tr>
<td>James R. Segedy, Gautam Biswas and Brian Sulcer</td>
<td></td>
</tr>
<tr>
<td>Learning Preferences and Motivation of Different Ability Students for Social-Competition or Self-Competition</td>
<td>283–293</td>
</tr>
<tr>
<td>Zhi-Hong Chen</td>
<td></td>
</tr>
<tr>
<td>Role Blending in a Learning Environment Supports Facilitation in a Robotics Class</td>
<td>294–306</td>
</tr>
<tr>
<td>Ilkka Jormakainen and Erkki Sutinen</td>
<td></td>
</tr>
<tr>
<td>Towards an Enhanced Learning Management System for Blended Learning in Higher Education Incorporating Distinct Learners’ Profiles</td>
<td>307–319</td>
</tr>
<tr>
<td>Sofia Balala Dias and José Alves Diniz</td>
<td></td>
</tr>
<tr>
<td>The Effects of the E-Book System with the Reading Guidance and the Annotation Map on the Reading Performance of College Students</td>
<td>320–331</td>
</tr>
<tr>
<td>Liang-Yi Li, Cheng-Yu Fan, Dong-Wei Huang and Gwo-Dong Chen</td>
<td></td>
</tr>
<tr>
<td>Potential Negative Effects of Mobile Learning on Students’ Learning Achievement and Cognitive Load—A Format Assessment Perspective</td>
<td>332–344</td>
</tr>
<tr>
<td>Hui-Chun Chu</td>
<td></td>
</tr>
<tr>
<td>Book Reviews</td>
<td></td>
</tr>
<tr>
<td>Simulation and Learning: A Model-Centered Approach</td>
<td>345–346</td>
</tr>
<tr>
<td>Reviewer: Dirk Ifenthaler</td>
<td></td>
</tr>
</tbody>
</table>
Guest Editorial: Game Based Learning for 21st Century Transferable Skills: Challenges and Opportunities

Francesco Bellotti1*, Rosa Maria Bottino2, Baltazar Fernández-Manjón3 and Rob Nadolski4
1DITEN, University of Genoa, Italy // 2Institute of Educational technology, National Research Council, Italy // 3DISIA, Universidade Complutense de Madrid, Spain // 4Open University of the Netherlands, Centre for Learning Sciences and Technologies, The Netherlands // franz@elios.unige.it // bottino@itd.cnr.it // balta@fdi.ucm.es // Rob.Nadolski@ou.nl

*Corresponding author

Digital serious games (SGs) (Gee, 2003; Prensky, 2003) offer a high potential to foster and support learning in educational and training settings. SGs aim at improving learning processes by providing attractive, motivating and effective tools. So far, effectiveness of SGs has been shown by recent studies (e.g., Connolly et al., 2012; Wouters et al., 2013), but the potential of SGs in education is still far to be fulfilled, in particular concerning higher-order learning goals (Connolly et al., 2012) and there is a growing need for educational technology research in this field. Moreover, education, cognitive and engineering methods and tools are needed for efficiently building and evaluating games as means that can provide effective learning experiences (Marfisi-Schottman, Labat & Carron, 2013; Bellotti et al., 2012).

This special issue focuses on analysing how digital SGs can contribute to the knowledge society’s higher demand towards acquiring transferable, transversal skills, that can be applied in different contexts, dealing with various scientific disciplines and subjects. Examples of such skills, often referred to as 21st century transferable skills, include, for example, collaboration, critical thinking, creative thinking, problem solving, reasoning abilities, learning to learn, decision taking, digital literacy (Voogt & Pareja Roblin, 2010).

This special issue explores particularly the challenges and opportunities presented by the use of digital SGs in formal learning contexts. The idea is to look not only to the tools (namely, the SGs) but also to the definition of meaningful practices through which such tools can be used effectively to reach specific learning goals (Bottino, Ott & Tavella, 2011). This means considering the whole learning environments in which games are integrated (including curricular/training goals, tools, tasks, methodologies, assumed roles and the context of use), which is in line with current research studies in technology enhanced learning, where technology design and use is increasingly considered in relation to the whole teaching and learning process.

This special issue was conceived in order to provide significant insights from the latest research work and to stimulate a fruitful dialogue between researchers engaged along the joint perspectives of educational SG design and use. The topic met a great interest from authors, as the call for papers received 50+ papers, out of which the guest editors, in collaboration with 113 reviewers, selected the following five papers.

The paper by Cowley et al. reports on a laboratory experiment combining evaluation methods from the fields of subjective learning assessment and of psychophysiology, considering various neurophysiological signals. The study identifies a relationship between learning outcomes and physiological measurements of mental workload, which opens new perspectives for SG user assessment.

The paper by DiCerbo presents an evidence model for assessing persistence, which is an important skill, in particular for goal achievement. Evidence extracted from log files of a commercial children game was used to identify players’ goals and to create a measure of persistence toward those goals. The results support the argument for a game-based measure of persistence.

Two papers (by Shah and Foster and by Eseryel et al.) concern pedagogical models ad-hoc developed to support effective use of SGs in formal education settings. Shah and Foster explored the ecological conditions necessary for implementing a system-thinking course in a 5th and a 6th grade classroom using a well established commercial game. The teacher successfully adopted a teaching model for game-based learning, and students showed statistically significant knowledge gains.
The paper by Eseryel et al. presents a theoretical model and describe an empirical investigation aimed at examining the interplay between learners’ motivation, engagement, and complex problem-solving outcomes in game-based learning. Findings suggest that learners’ motivation determine engagement, which in turn determines development of complex problem-solving competencies.

A last paper, by Di Blas and Paolini, presents the outcomes of a large case-study of four formal education programs exploiting serious games based on multiuser virtual environments. The programs proved to be highly effective in fostering a number of transversal skills - in particular collaboration.

The guest editors are proud of presenting a balanced mix of papers, especially in terms of perspectives and addressed topics. The special issue shows that a large amount of work is being done in order to develop models and methods for effective serious game deployment, especially in formal education contexts. We believe that more in-depth analysis and extensive/comparative user studies are necessary for a better validation of serious game effectiveness, also concerning 21st century skills, and for understanding when and how to use games to complement other educational means and approaches. This will be key to developing a new generation of serious games that improving aspects such as assessment (Bellotti et al., 2013), feedback (Hays, Lane, & Auerbach, 2013), analytics (del Blanco et al., 2013) and collaboration support (Hummel et al., 2011) - should lead to increasing effectiveness in terms of content presentation/adaptation and user motivation and coaching.

References


Learning When Serious: Psychophysiological Evaluation of a Technology-Enhanced Learning Game

Ben Cowley1,2*, Martino Fantato1, Charlene Jennett3, Martin Ruskov3 and Niklas Ravaja1,4,5
1School of Business, Aalto University, 00076-AALTO, Finland // 2Cognitive Science Unit, University of Helsinki, P. O. Box-9, 00141, Finland // 3Department of Computer Science, University College London, WC1E 6BT, UK // 4Department of Social Research, University of Helsinki, P. O. Box-54, 00141 University of Helsinki, Finland // 5Helsinki Institute for Information Technology HIIT, P. O. Box-15600, 00076-AALTO, Finland
ben.cowley@helsinki.fi // martino.fantato@aalto.fi // c.jennett@cs.ucl.ac.uk // m.ruskov@cs.ucl.ac.uk // niklas.ravaja@aalto.fi

*Corresponding author

ABSTRACT
We report an evaluation study for a novel learning platform, motivated by the growing need for methods to do assessment of serious game efficacy. The study was a laboratory experiment combining evaluation methods from the fields of learning assessment and psychophysiology. 15 participants used the TARGET game platform for 25 minutes, while the bio-signals electrocardiography, electrodermal activity and facial electromyography were recorded. Learning was scored using pre- and post-test question-based assessments. Repeated-measures analysis with Generalised Estimating Equations was used to predict scores by tonic psychophysiological data. Results indicate some learning effect, plus a relationship between mental workload (indexed by electrocardiography) and learning. Notably, the game format itself influences the nature of this relationship. We conclude that a high quality of insight is afforded by the combination of subjective self-report and objective psychophysiology, satisfying two of three observable domains.

Keywords
Technology enhanced learning, Serious games, Heart-rate variability, Mental workload, Psychophysiology, Evaluation, Competence development

Introduction
In today’s global market, human capital is a recognized strategic asset in companies. Learning and training play a foundational role in talent management, but establishing effective learning strategies across enterprises remains a costly challenge without measurable return.

To support the effort to build such strategies, we report on an evaluation of a game-based Technology-Enhanced Learning (TEL) platform, which teaches about soft skills using project management scenario simulations. Based on positive outcomes from a similar previous study (Cowley, Ravaja, & Heikura, 2013), we designed a combined-methods approach to the evaluation. A multi-trial protocol allowed a repeated measures self-report battery alongside the measurement of physiological signals to index psychological constructs. The two methods combined are a complementary data-gathering tool, because self-report is subjective, discrete and explicit while psychophysiology is objective, continuous and implicit. Thus the study was carried out to examine whether psychophysiological recordings obtained during serious game play predict (short-term) learning outcomes as measured by a pre- and post-test assessment tool. We found, among others, a relationship between mental workload and learning. This has important implications for researchers interested in measuring learning (unobtrusively) in future research.

The experiment used a within subjects design based on adjustment of the participant’s knowledge of project management via the learning platform. The independent variables were the physiological effects on the player of exposure to source of topic-relevant education (namely the game), while the dependent variable was the knowledge of the participant. Learning was assessed using questionnaires of two types: one set applied before and after play to test learning performance, and one set of self-report questionnaires to establish the ‘felt’ experience of the participant.

We begin by describing the state of the art underpinning this experiment. We cover experiment methodology and then detail our results under three themes: psychophysiological predictors of learning; test-based assessment of learning; subjective mood self-report. Finally we offer our discussion and conclusion.
State of the art

Educational game efficacy has been well debated (Egenfeldt-Nielsen, 2006; Gee, 2006). McQuiggan, Lee, & Lester (2006) draw a parallel between the factors describing student engagement and those involved in game play. However, learning does not necessarily follow engagement, as argued by Kirschner, Sweller, & Clark (2006) who point out that discovery, problem-based, experiential and enquiry-based techniques are the main tools of games, but all require prior knowledge on the part of the student to evoke learning. Some suggest the solution is in scaffolding the game, i.e., instructional support during learning to reduce cognitive load (O’Neil, Wainess, & Baker, 2005).

Given recent positive results (Blunt, 2009; Ritterfeld, 2009) – and bearing in mind that some form of learning is almost always part of play (Koster, 2005) – the relevant question becomes: how will a given game work? Will a particular game teach retained, transferable skills which are the ones intended by the designers, or will it teach skills only valuable within the game context? This has become a strong theme in serious games research (Guillén-Nieto & Aleson-Carbonell, 2012).

When examining how such design variants actually work ‘in the field’, the players’ psychological and physiological experience is of central interest, motivating the need to objectively measure this subjectivity. For the assessment of subjective experience, e.g. emotions, there are three observable domains (Bradley & Lang, 2000): i) a subjective cognitive experience (e.g. assessed by questionnaires), ii) behavioural expressions (i.e. actions and behavioural patterns assessed by implicit techniques) and iii) psychophysiological patterns. Three features of the physiology which are particularly interesting in respect of learning are arousal, cardiac and facial musculature – i.e. bodily activation, regulation and pleasure or displeasure in response to the experimental activity. These may act as a sufficient causal explanation (Peters, 1960, p-11) of observed learning.

The basic premise of psychophysiological methods for evaluation is that the study participant/player cannot give inaccurate physical signals (discounting acquisition issues), and the acquisition of signals is non-intrusive, freeing the user/player’s attention. Psychophysiological methods are particularly useful for objectively examining experience: because the physiological processes measured are mostly non-voluntary, the measurements are not contaminated by participant answering style, social desirability, interpretations of questionnaire item wording, or limits of their memory (Ravaja, 2004).

There is emerging evidence suggesting that the synchronization of activity of different physiological response systems (i.e. response coupling) may reflect the central state of the individual particularly well (Chanel, Rebetez, Bétrancourt, & Pun, 2011), prompting our use of three separate biosignals: electrodermal activity (EDA), electromyography (EMG) and electrocardiography (ECG).

Several studies have shown that digital games (i.e., an active coping task) elicit considerable emotional arousal- or stress-related cardiovascular reactivity in terms of heart rate (HR) and blood pressure (Johnston, Anastasiades, & Wood, 1990). Previously found convergent relations between HR and arousal during digital game playing suggest that HR covaries primarily with emotional arousal during playing (Ravaja, Saari, Salminen, Laarni, & Kallinen, 2006). Henelius, Hirvonen, Holm, Korpela, and Muller (2009) explored the ability of different short-term heart rate variability (HRV) metrics to classify the level of mental workload during a variable task-load computerized test, with good results. Additionally Nishimura, Murai and Hayashi (2011) found that mental workload was well indexed by HRV in a simulator learning task; Hercegfi (2011) reported on a method to assess mental effort of human-computer interaction from HRV and other signals, reporting improvement over existing approaches.

Plotnikov et al., (2012) and Nacke, Stellmach, and Lindley, (2010) have each described approaches to measuring the engagement of players using electroencephalography.

Facial Electromyography (fEMG), direct measurement of the electrical activity associated with facial muscle contractions (Tassinary & Cacioppo, 2000), has the potential to record the unconscious emotional reactions of the player to interaction with in-game stimuli. Recording at the sites of Zygomaticus major (cheek) and Corrugator Supercili (brow) can index positive and negative valence (Lang, Greenwald, Bradley, & Hamm, 1993; Ravaja, Saari, Turpeinen, et al., 2006; Witvliet & Vrana, 1995). Recording at the Orbicularis Oculi (periocular) can index high arousal positive valence (Ekman, Davidson, & Friesen, 1990).
Arousal is most often measured with EDA (or skin conductance level; also sometimes called galvanic skin response) (Bradley, 2000; Dawson, Schell, & Filion, 2000). EDA is an often-used physiological measure when studying digital gaming experiences (e.g., Mandryk & Atkins, 2007), as it is less susceptible to misinterpretations compared to facial EMG. The neural control of eccrine sweat glands—the basis of EDA—predominantly belongs to the sympathetic nervous system that non-consciously regulates the mobilization of the human body for action (Dawson et al., 2000).

Our primary evaluation approach was to examine the relationship of the psychophysiological signals with learning questionnaires, self-reported game experience and game experience questionnaires. We use tonic values of psychophysiological signals to index various cognitive and emotional processes which can contribute to learning. This approach is novel in its domain, as few psychophysiological studies focus on this area of soft-skills game-based learning (GBL); it is also novel in terms of methodology, as prior studies of GBL mainly used event-related analysis of the biosignals; however because learning in this type of game happens over long time periods, with players who can construct concepts from non-linear relationships in the data they are presented with, a tonic analysis approach is more appropriate.

Methodology

Participants

We enrolled 15 right-handed participants (seven females), randomly sampled from respondents (self-selected volunteers) to advertisements. Their ages ranged from 21-33 years, the mean age being 25.87 years (SD = 3.85). Participants were rewarded with three cinema tickets each. Regarding their background, all 15 participants were novices to project management. 12 from 15 used IT to support their learning, including activities such as online courses, web research, reading online journals, using mind-map programs, watching video lectures. Four out of 15 had used role-play to support their learning: three took part in a course that involved role-play and discussion; one enjoyed ‘larping’ (live action role-playing). 10 of 15 currently played computer games of some kind. Two said that they gave up playing such games due to lack of time. Three did not play games.

TARGET platform

The EU-funded Transformative, Adaptive, Responsive and enGaging EnvironmenT (TARGET) project has developed an innovative TEL environment focused on project management. It was an ideal test case for our current work, because universities/enterprise settings that wish to use the TARGET platform might want to measure whether their employees/students are learning without disturbing their playing experience.

The TARGET platform features learning scenarios that employ interactive storytelling and evidence-based performance assessment. In the Stakeholder Management (SM) scenario (see Figure 1) learners get to play the role of the project manager Ingrid. Ingrid is responsible for developing a windmill electricity farm project, and has access to project management tools such as a Gantt chart which are used during the scenario to scaffold the task simulation. Ingrid’s task is to negotiate with Jens, a local farmer, to convince him to sell his land for building an access road. Negotiation happens in the form of unstructured dialogue; learners converse via a text field. To help ensure convergence in how didactic ideas are expressed, the dialogue input field provided suggestions for sentence auto-completion. Participants were advised to use these as much as possible. Learners familiar with the scenario and interface usually take up to 10 minutes to negotiate, regardless of success. To ensure such familiarity participants were given a training session.

After the negotiation experience, learners are presented with the evidence-based ‘Competence Performance Assessment’ module (CPA). The CPA includes a video re-play of the experience synchronised to a line-graph showing the learner’s estimated performance in four competences: negotiation, communication, trust-building, and risk and opportunity (see Figure 2). The playback gives learners a chance to review their behaviour, relate their actions to their dynamic competence assessment, and interpret where they went wrong or right. The CPA module is described in detail in (Parodi et al., in press).
There are several ways that the TARGET platform aims to achieve learning:

1. By playing the scenario and observing the causal logic of the dialogue system, the learner can see how factors such as style of approach, mood, small talk, etc. influence the negotiation outcome.

2. By reflecting on their performance using the CPA module, they can review their behaviour and think about how they would carry out the negotiation differently next time.

3. The learner is able to replay the scenario multiple times (in our study we instruct learners to play it twice). This allows the learner to try out different strategies, and through repeated play he/she can learn the operative principles for successful negotiation in the scenario.
Psychophysiological measures

We recorded continuously three sets of electrical potentials: EDA is a measure of arousal; EMG when applied facially (fEMG) helps index emotional expression; and ECG measures the heart’s potential field and can help index many processes such as mental workload (Cowley et al., 2013).

fEMG was recorded at three separate sites: the muscle regions Corrugator Supercilii (CS, above the brow, indexes negative emotion), Zygomaticus Major (ZM, on the cheek, indexes positive emotion) and Orbicularis Oculi (OO, below the eye, indexes sincere positive emotion). Electrodes were filled with Synapse conductive electrode cream (Med-Tek/Synapse, Arcadia, CA), and impedances were ensured to be below 10kΩ.

To record EDA, electrodes were placed to onto the proximal phalanges of the non-dominant hand. Electrodes were Ag/AgCl filled with TD-246 skin conductance electrode paste (Med Assoc. Inc.). ECG was recorded at the manubrium, lower left rib and neck using electrodes with stud connector 35x45 mm (Spes Medica S.r.l.). For the recordings, the VarioPort ARM was used as DC amplifier and recording device. The specifications of the amplifier can be found in Table 1.

<table>
<thead>
<tr>
<th>Signal</th>
<th>Amplification factor</th>
<th>Input range</th>
<th>Resolution (with 16 bit converter)</th>
<th>Frequency Range</th>
<th>Sampling Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECG</td>
<td>214 (±2%)</td>
<td>±5.8mV</td>
<td>0.18µV/bit</td>
<td>0.9-70Hz</td>
<td>64Hz</td>
</tr>
<tr>
<td>EMG</td>
<td>4899 (±2%)</td>
<td>±255mV</td>
<td>0.008µV/bit</td>
<td>57-390Hz (±2%)</td>
<td>128Hz</td>
</tr>
<tr>
<td>EDA</td>
<td></td>
<td>0-70µS</td>
<td>0.001µS</td>
<td></td>
<td>32Hz</td>
</tr>
</tbody>
</table>

Learning assessment materials

Several learning assessments were designed to test for learning, see Table 2. An 8-item multiple choice questionnaire (MCQ) was administered pre- and post-TARGET. Each question was presented with 5 possible answers. The MCQ was created with the aid of an instructor that used the SM scenario as a teaching exercise with his project management students. The questions were designed to cover key points of negotiation, communication and trust-building.

A 10-item self-assessment of learning questionnaire (SAL) was administered post-TARGET only. For each item, participants are asked to rate their level of agreement on a scale from 1 (strongly disagree) to 7 (strongly agree). Four items refer to competence performance (SAL-performance) and four items refer to competence learning (SAL-learning).

<table>
<thead>
<tr>
<th>Learning Materials</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCQ</td>
<td>1. What is a Gantt chart used for?</td>
</tr>
<tr>
<td></td>
<td>2. What is usually the best way to start a negotiation with a stakeholder?</td>
</tr>
<tr>
<td></td>
<td>3. When you are selling a project to the stakeholder, which of the following should be your priority to address?</td>
</tr>
<tr>
<td></td>
<td>4. When you want to get commitment from the stakeholder, which is the better approach and why?</td>
</tr>
<tr>
<td></td>
<td>5. Why is building trust with a stakeholder important?</td>
</tr>
<tr>
<td></td>
<td>6. Which of the following is a way to break trust with a stakeholder?</td>
</tr>
<tr>
<td></td>
<td>7. Ideally what is the best way to communicate with a stakeholder?</td>
</tr>
<tr>
<td></td>
<td>8. What language should you use when communicating with a stakeholder?</td>
</tr>
</tbody>
</table>
SAL  

**Experience:**
1. Did you find the Gantt chart useful for managing your project?
2. Did you enjoy the experience of negotiating with the game character(s)?

**Competence performance:**
- Do you think you performed well for
  3. negotiation?
  4. trust-building?
  5. communication?
  6. risk and opportunity?

**Competence learning:**
- Did playing the game and reflecting on your experience help you to learn about
  7. negotiation?
  8. trust-building?
  9. communication?
  10. risk and opportunity?

---

**Additional materials**

We used the self-report item Game Experience Questionnaire (GEQ). The GEQ measures subjective experiences relevant to play including Competence, Sensory & Imaginative Immersion, Flow, Tension, Challenge, Negative affect and Positive affect.

**Evaluation design**

The evaluation lasted from 2h 10min to 3h 30min (M = 2h 47min, SD = 30min), including preparation and answering questionnaires. The protocol of experiment followed 15 stages:

1. Welcome and briefing (~7min). Participants were informed of the general nature of the evaluation and asked their background information.
2. Pre-test learning measures (~35min). The second stage was to answer the pre-test learning questionnaire, wherein there were no timing restrictions; participants were then trained how to use the platform.
3. TARGET training (~10min). Training material explained the main features and operational details of the TARGET platform.
4. Electrodes and amplifier attached (~30min). After attaching the electrodes and testing the signals, an initial 5-min resting time baseline preceded the playing session of 10 minutes.
5. Baseline (5min). Both baseline and play session times were fixed.
6. Play TARGET game first time (10min). Participants played twice with the same procedure.
7. Self-report: GEQ (~2min). At the end of both game sessions participants had to fill the same self-report survey to assess subjective mood.
8. CPA reflection (5min). In between of the two sessions, there was a 5-minute reflection period, where participants were asked to review their experience using the CPA module and to think about what they did well, what they did poorly, and what they would do differently next time.
9. Break (5min). The short break was used to equalise the physical state of the player between play sessions, and check signals.
10. Baseline (5min). Due to the potential change of physical state after playing, a new baseline was recorded.
11. Play TARGET game second time (10min).
13. Electrodes off (~5min).
14. Post-test learning measures (~30min). Finally, participants had to answer the post-test learning questionnaire, which was largely the same as the pre-test.
15. Debriefing (~5min).
Analysis

Learning in the TARGET game was assessed by two pre- post-questionnaire instruments, and one self-assessment, as described above. Four dependent variables (DV$s) were derived from these by the following equalities.

- $MCQ_{\text{diff}} = \text{MCQ scores, post-test minus pre-test}$.
- $EXP = \text{average scores of SAL experience}$.
- $Perform = \text{average scores of SAL competence performance}$.
- $Learn = \text{average scores of SAL competence learning}$.

Psychophysiological data processing procedure

EDA signal was pre-processed using Ledalab (v 3.43) in batch mode: down-sampled to 16 Hz and filtered using Butterworth low-pass filter with cut-off 5 Hz and order 8. The signal was divided in phasic (EDAP) and tonic (EDAT) components using the nonnegative de-convolution method (Benedek & Kaernbach, 2010). Thus phasic EDA and tonic EDA were two of our Independent Variables (IV$s$); IV$s$ were defined as mean values of the relevant signal over some epoch – in this experiment, epoch length was one minute. Using phasic EDA does not contradict our analysis approach because it is nevertheless a continuous signal component which is hypothesised to be the response to discrete events.

ECG signal was pre-processed using Ecglab toolbox for MATLAB (Carvalho, Da Rocha, Nascimento, Souza Neto, & Junqueira Jr, 2002). R-peaks were identified from the original 64 Hz series and corrected for ectopic beats. Inter-beat interval (IBI) time series was obtained by interpolating with cubic splines at 4 Hz. We extracted two features as IV$s$: Heart Rate (HR) and Heart Rate Variability (HRV).

EMG signals were pre-processed using standard MATLAB functions. They were filtered using 50Hz notch filter, smoothed and square roots of the means squared features were extracted. The resultant signal for each fEMG electrode location was an IV – that is, fEMG 1 = Zygomaticus major, fEMG 2 = Corrugator Supercilii, fEMG 3 = Orbicularis Oculi.

In total there were seven IV$s$ representing the recorded psychophysiology. Every psychophysiological variable also had a baseline, which was derived by the same pre-processing procedure as the mean value over the entire baseline recording period. Baselines are important to correct psychophysiological data which does not conform to absolute ranges – one person, on any given day, may present with greater or lesser baseline levels of activation of any given psychophysiological signal. From all signal files, one minute mean values of each of the IV$s$ signals described above were extracted and tabulated with the background and self-report data gathered from participants, into a 2D data matrix suitable for analysis, oriented with repeated measures over time row-wise and variables/factors column-wise.

Statistical analysis

The full list of DV$s$ was $MCQ_{\text{diff}}, EXP, Perform$ and $Learn$. The full list of IV$s$ included: EDAP, EDAT, HR, HRV, fEMG Zygomaticus, fEMG Orbicularis Oculi, fEMG Corrugator Supercilii. Additional factors included Gender and the dichotomous background variables numbered 2-4 in the list above (first paragraph in Methods section): IT learner, role-player and game-player. Covariates were the baseline measures for each of the IV$s$, and Age.

The Generalized Estimating Equations (GEE) procedure in SPSS was used to analyse the data, and separate analyses were carried out for each of the DV$s$. For every model, we specified participant ID as the subject variable, and gameplay trial and minute as the within-subject variables. On the basis of the Quasi-likelihood under Independence Model Criterion (QIC), we specified Independent as the structure of the working correlation matrix. Due to the ordinal nature of the DV$s$, we specified an ordinal distribution with the Logistic link function.

GEEs are an extension of the generalized linear model, and were first introduced by Liang and Zeger (Liang & Zeger, 1986). GEEs allow relaxation of many of the assumptions of traditional regression methods such as normality and homoscedasticity, and provide the unbiased estimation of population-averaged regression coefficients despite possible misspecification of the correlation structure.
Results

Psychophysiological predictors of learning

Taken from the complete analysis, the statistically significant psychophysical results are summarised in Table 3, showing each physiological variable (IV) that predicted learning (DV) and associated statistics. Analysis of participants’ background data showed some important group-level distinctions. A t-test of gender across the five DVs mentioned in the previous section showed that men and women scored significantly differently in MCQdiff, mean male scores were higher, $t_{(13)} = 2.84, p < .05$. Thus we know that men performed significantly better than women on MCQdiff, but the small sample size gives little chance of drawing a conclusion from this alone. A t-test of levels of role-playing across the Perform DV showed that those who said they used it in learning scored significantly higher, $t_{(14)} = 2.2, p < .05$. These results give context to the analysis, and indicate which of the available data to include as factors or covariates in the GEE models.

Table 3. Summarised statistical results for psychophysiology. Only significant results are reported

<table>
<thead>
<tr>
<th>IV</th>
<th>B</th>
<th>SE</th>
<th>Wald Chi-Square (df=1)</th>
<th>DV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task-level HRV ***</td>
<td>-0.59</td>
<td>0.16</td>
<td>12.87</td>
<td>MCQdiff</td>
</tr>
<tr>
<td>Tonic EDA *</td>
<td>5.03</td>
<td>2.26</td>
<td>4.95</td>
<td>MCQdiff</td>
</tr>
<tr>
<td>fEMG Zygomaticus ***</td>
<td>-1.54</td>
<td>0.53</td>
<td>8.49</td>
<td>MCQdiff</td>
</tr>
<tr>
<td>fEMG Orbicularis *</td>
<td>-0.85</td>
<td>0.42</td>
<td>4.03</td>
<td>MCQdiff</td>
</tr>
<tr>
<td>HR **</td>
<td>3.19</td>
<td>1.18</td>
<td>7.32</td>
<td>Perform</td>
</tr>
<tr>
<td>HRV *</td>
<td>0.49</td>
<td>0.25</td>
<td>3.70</td>
<td>Perform</td>
</tr>
<tr>
<td>EDA ***</td>
<td>-2.4</td>
<td>0.85</td>
<td>8.05</td>
<td>Learn</td>
</tr>
</tbody>
</table>

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$

GEE models were tested for each of the five learning measurement DVs, with main effects of one IV (out of seven), the baseline (for that particular IV), Age, Gender, and the dichotomous background variables IT learner, role-player and game-player. This gave 35 models. Four of the seven IVs for MCQdiff were significant.

Task-level HRV was negatively associated with MCQdiff. That is, decreased HRV during game playing was associated with better learning. In contrast tonic EDA was positively associated with MCQdiff. Thus, general level of basal arousal was increased for those with better learning.

Both fEMG Zygomaticus and fEMG Orbicularis are indices of positive emotion, as explained above. fEMG Zygomaticus was negatively associated with MCQdiff, and fEMG Orbicularis was negatively associated with MCQdiff. Thus, positive emotional responses were decreased for those who showed better learning by multiple choice questions. Any link between MCQdiff and a negative or neutral emotional reaction to the game is not suggested by statistical testing of participants’ GEQ self-reports. Two items tested as significant, but only marginally; they were GEQ responses ‘I forgot everything around me’ (GEQ5) and ‘I felt challenged’ (GEQ12), by SPSS General Linear Model (GLM) Univariate procedure with MCQdiff as DV and each self-report item as covariate.

HR was positively associated with Perform. HRV was only marginally significant, but was positively associated with Perform. Therefore, in the Perform-related self-assessments participants with higher HR and HRV rated their performance more highly.

Finally, phasic EDA was negatively associated with Learn. This implies that the more participants were activated/aroused by the events of the game, the lower that they rated themselves in the Learn-related self-assessments. All results are further explored and discussed in the Discussion below.

Question-based assessment of learning

Multiple choice questionnaire (MCQ)

The mean MCQ score before TARGET was 4.67 out of 8 (SD = 1.50). The mean MCQ score after TARGET was 5.20 out of 8 (SD = 1.47). A paired samples t-test revealed that this was not a significant difference, $t_{(14)} = -1.372, ns.$
Figure 3 shows frequencies of participants that answered each MCQ correctly before and after TARGET. The number of participants that responded after TARGET with the correct answer increased for the MCQs 1, 2, 3, 6, and 7; and decreased for the rest. Paired samples t-tests revealed that there were significant differences for MCQ1 ($t_{(14)} = -2.449, p < .05$) and MCQ2 ($t_{(14)} = -3.055, p < .01$). This suggests that after playing TARGET, participants had better knowledge of Gantt charts (MCQ1) and the importance of small talk for negotiations (MCQ2). Additionally, for those items which did not have individually significant gain scores, we tested the items with positive gains (3, 6 and 7) against those with negative gains (4, 5, and 8). A paired-samples $t$-test for summed gain scores showed significant difference, $t_{(14)} = 2.703, p < .05$.

![Figure 3. Response frequency and t-tests for pre-to-post MCQs](image)

**Self-Assessment of learning questionnaire (SAL)**

In Figure 4 we present frequencies of responses, where ratings of 1-3 are grouped under “disagree,” 4 as “neutral,” and 5-7 are grouped under “agree.” To specify with respect to the DVs, SAL items 1 and 2 relate to experience; SAL items 3, 5, 7 and 9 relate to competence performance (DV Perform); SAL items 4, 6, 8, and 10 relate to competence learning (DV Learn). The results suggest that opinions were mixed for experience, competence performance and competence learning.

Although SAL was not a test, we can still gain some insights from the relative distribution of responses. For instance, we can see that the greatest proportion of positive responses, six, were given to items 3 and 4: participants believed they both performed well and learned about negotiation. In contrast the least number of positive responses, two, were given to items 7 and 9: performed well for communication and for risk and opportunity. These two items also had low disagreement rates – therefore mainly there was a large neutral response. Most negative responses, eight, went to item 10: learning about risk and opportunity. Items with the least negative responses were 2, 3 and 7: enjoyed negotiating, performed well for negotiation and communication.

![Figure 4. Sample response statistics for the SAL questions](image)
State-based self-reports

The self-report items asked after each game session give a subjective impression of the game experience, complementing the psychophysiological data. Mean and standard deviation is given for each scale item, and figures 4-6 describe the responses in terms of the percentage of respondents corresponding to the low-medium-high levels in the scale (summed over both response times, but missing one participants’ data, i.e., 29).

Game experience questionnaire (GEQ)

Participants rated their game experience from 1 to 5, with 14 items which load to seven factors: Competence (M = 2.12, SD = 0.07), Sensory & Imaginative Immersion (M = 2.48, SD = 0.83), Flow (M = 2.22, SD = 0.22), Tension (M = 2.76, SD = 0.29), Challenge (M = 2.59, SD = 0.44), Negative affect (M = 1.97, SD = 0.00), Positive affect (M = 2.69, SD = 0.15).

In Figure 5 we present the response distribution. We used a t-test to determine which response distributions differed significantly from the response ‘Not at all’; which we interpret to mean which experiential areas the game elicited some reaction. By this test we can say that a significant (all at \( p < .05 \)) number of participant responses were either moderately or extremely positive for items Positive affect, \( t_{(14)} = 2.9 \); Tension, \( t_{(14)} = 2.3 \); Sensory & Imaginative Immersion, \( t_{(14)} = 2.4 \), and Challenge, \( t_{(14)} = 2.8 \); contrasting with items Negative affect, Competence, and Flow, for which two-thirds or more responded ‘not at all’.

Discussion

The study aimed to evaluate a novel TEL game, using a combined methods approach with learning assessment and psychophysiology. We propose that the combined approach added value by giving insights where each method alone suffers from ambiguity of interpretation.

Based on MCQ pre- to post-assessment, we conclude that the TARGET game has potential to support learning, due to two positive MCQdiff results. However because the gain scores were not significant overall, it seems the platform is not yet operating at full potential. We will attempt to explain why using the measurements taken.

The two significant MCQdiff results, from MCQs 1-2, contrast in their content with MCQs 2-8, which showed no significant improvement. Specifically, MCQ 1, “what is a Gantt chart...” deals with a more concrete issue reflecting a defined action that the player had to take in the game. For MCQ 2, “the best way to start a negotiation...” is more often realised than the issues from later MCQs because participants start negotiations more often than they successfully conclude them.

Contrasting those MCQs which had positive gains against those with negative gains, MCQs 3, 6 and 7 seem to be more straightforward, dealing with simpler concepts. They may also have been easier to remember from the pre-test, resulting in a gain based on priming.
The SAL results, i.e., areas where participants felt that they learned and performed well, imply that they may have had difficulty parsing causal mechanics of game. In other words, they could see that the outcomes of their negotiations were often good, but they did not have much access to the reasoning behind success, i.e., the risks and opportunities underlying their project were not well understood.

For GEQ affective self-report, there is an apparent paradox that relatively good Positive affect and Immersion can occur in the presence of relatively high tension and low competence. We can shed some light on the cause of this puzzle by studying responses for one particular factor. In this (short-form) version of GEQ, Immersion is based on two items: 1 “I was interested in the game’s story,” and 2 “I found it impressive.” For item 1, only 28% responded 'Not at all' (the lowest for all responses); for item 2, this figure was 69%. Thus we see a dichotomy between participants’ interest in, and how impressed they were by the game: perhaps indicating that problems were more in the delivery than the content. It appears the quality of the experience was not rewarding, evidenced by low perceptions of success (69% ‘Not at all’ responses to GEQ:Competence). Oddly, this contrasts with low perceptions of challenge (50% ‘Not at all’ responses for GEQ:Challenge).

Overall, self-report and testing results show that participants felt intrigued and challenged, but also frustrated by their inability to win the game. These results may suggest the interpretation that difficulty in using the game impeded important learning: however the standalone value of self-report evidence is the subject of some debate (Yannakakis & Hallam, 2011), and psychophysiology offers a complementary perspective that helps address the specific concerns of subjectivity and imprecision.

Psychophysiology

Examining the various IVs which predict MCQdiff reveals the nature of the relationship between player and game. Recalling that MCQdiff is a gain score, similarly to (Cowley et al., 2013), we can draw on the picture established there with respect to HRV. It is evident that learning is not immediately dependent on the particular direction of the HRV relationship. Rather, because HRV indexes mental workload, a given game may require a) more or b) less mental workload relative to baseline values in order to induce learning. The TARGET game clearly falls in category a), perhaps because it is designed as a business simulator. In our proxy game study (Cowley et al., 2013), we used the educational game Peacemaker (Impact Games 2007). The Peacemaker serious game was the designed to teach a peace-oriented perspective on the Israel-Palestine conflict. It is a point-and-click strategy game, where the player acts as a regional leader and must choose how to react to the (deteriorating) situation, deploying more or less peaceful options from a menu in a turn-based manner. It is possible that players who performed better in Peacemaker took advantage of cognitive efficiency enabled by the more game-like and polished elements of that game’s design, and thus Peacemaker falls in category b).

For higher scoring players of the TARGET game, the fact that decreased HRV indicates increased mental workload is corroborated by their increased tonic EDA. Basal arousal is increased as a function of the task-related workload.

It seems that players who learned may not have found the experience unpleasant, but it was nevertheless not pleasant, was challenging and possibly frustrating, and took effort. This is contextualised by the results for the SAL questions, where participants who were more aroused in response to the events of the game also rated themselves more critically on their performance and learning. If the trend among participants was to perform better with less positive, more negative emotional responding, then it follows that activation in response to game events would have been generally of a negative nature. In other words, failure/setbacks (being more common) aroused a response more often than success.

In general, because four of seven IVs for MCQdiff and four of seven items from GEQ were significant, we might conclude that the game was more affecting than not, creating an intriguing if not perfect experience. Combining self-report, testing and psychophysiology, the results suggest a challenging experience for participants as they attempted to learn from the beta TARGET platform.
Issues and future work

Although the sample size may be unusually small in the educational literature, it is reasonable for psychophysiological studies. One can also claim based on a Bayesian inferential argument (Wagenmakers, 2007) that for a given \( p \) value, small \( N \) actually constitutes greater evidence against the null hypothesis.

Some study design issues became apparent on analysis. The dependent metrics used were all based on interstitial self-report, a method which has pros and cons. The reports come only at quite infrequent intervals. Using a regular pop-up prompt to gather closer-to-real-time data can increase the frequency, although only at the cost of invasiveness. In this study, using pop-up self-report was infeasible at the time and is a subject for future work.

Further study in the area may benefit from use of more psychophysiological signals such as EEG. While EEG itself indexes several informative features of player experience (Nacke et al., 2010), others have also found evidence that the fusion of several physiological modalities increases emotion recognition accuracy (Chanel et al., 2011) – offering the potential to apply such methods in TEL contexts. Such systems have been developed for the TARGET project (Bedek et al., 2011) and thus early work suggests their applicability also to learning games.

Conclusion

In this study we analysed psychophysiological recordings of players of the TARGET game with respect to their learning, as measured by pre- and post-test questionnaires. Several interesting results were found, pointing to a specific form of interaction for this game, where best results follow an effort of high mental workload and serious play. This may be opposed to types of TEL games where cognitive efficiency builds from the exigencies of playing and thus reduced mental workload is predictive of better learning (Cowley et al., 2013). In other words, given that HRV is negatively correlated with workload, for complex games a decrease in HRV can be a prediction of learning, while for simpler games an increase in HRV can be a prediction of learning. The difference may lie in the level of complexity which players are expected to master without some scaffolding from entertainment game-design techniques such as skill-chaining (building skills sets from easy component skills mastered one at a time). The closer that a TEL game gets to representing a real-world problem in real-world terms, the greater will be its inherent complexity and required mental workload.

The capability to make this distinguishing insight seems to be afforded by the combination of subjective self-report and objective psychophysiology, suggesting a methodology which can inform directions for future evaluation of such types of learning platforms. While interpretation of evaluation tests is always dependent on the researchers to some degree, the application of empirical techniques can help by corroborating or falsifying our stories.

Acknowledgments

This research was undertaken in the TARGET project, partially-funded by the European Community under the Seventh Framework Programme (Grant Agreement IST 231717).

References


14


Game-Based Assessment of Persistence

Kristen E. DiCerbo
Pearson, 400 Center Ridge Drive, Austin, TX 78753 // kristen.dicerbo@pearson.com

ABSTRACT
Interest in 21st century skills has brought concomitant interest in ways to teach and measure them. Games hold promise in these areas, but much of their potential has yet to be proven, and there are few examples of how to use the rich data from games to make inferences about players' knowledge, skills, and attributes. This article builds an evidence model for the assessment of persistence from Poptropica, a popular commercial game for children. Task persistence is an important skill related to successful school and work outcomes, particularly given new, complex tasks requiring sustained application of effort. Evidence extracted from log files of the game was used to identify players with a particular goal and then create a measure of persistence toward that goal. The results support the ability to create an assessment argument for a game-based measure of persistence.

Keywords
Game-based assessment, Persistence, Measurement models, Educational data mining

Introduction

The digital revolution is fundamentally changing the nature of work, the problems workers are asked to solve, and therefore the types of skills needed in today’s world. Today’s workers must be able to apply their knowledge to more complex tasks, generate creative solutions to multi-faceted problems, and navigate interwoven systems. Employers indicate that skills such as problem solving, communication, collaboration, and creativity, as well as personal attributes such as adaptability, persistence, and resilience are becoming more important, and have been labeled “21st century skills” (Casner-Lotto & Barrington, 2006; Fadel, 2011). However, employers also indicate that employees often lack these essential skills (American Management Association & Partnership for 21st Century Skills, 2010), leading to a push for them to be taught and assessed in schools.

However, these skills, by their very nature, do not lend themselves well to traditional methods of assessment. Most traditional tests present highly decontextualized individual items to learners. The 21st century skills and attributes of interest require application in context as part of complex tasks for accurate measurement. In addition, traditional assessment often interrupts the learning process in order to gather information and does little to motivate the learner to put forth effort, further jeopardizing our ability to gain valid estimates of skills and attributes. In response to these concerns, there has been growing interest in, and investigation of, the use of games to assess 21st century skills (Shaffer et al., 2009; Shute, 2011).

Games are attractive as assessment tools for a number of reasons. First, they allow us to make observations in contexts closer to those in the real world, creating the complex scenarios required to evaluate the application of knowledge and skills. Second, we know that games are engaging and motivating and that assessments are more valid when students are more motivated by them (Schmit & Ryan, 1992; Sundre & Wise, 2003). Third, we know that the vast majority of students (for example, 97% of teens aged 12-17 in the United States) already play digital games (Lenhart et al., 2008). This means we do not have to stop students’ daily activity in order to gather assessment information. Rather, we can tap into the digital ocean of data already being produced by their online activity to identify and accumulate evidence about what players know and can do from that (DiCerbo & Behrens, 2012). Finally, games and assessment share a similar process loop of activity presentation, activity completion, evidence identification, evidence accumulation, and presentation of the next activity (Behrens, Frezzo, Mislevy, Kroopnick, & Wise, 2006), making games ready targets for assessment efforts.

However, before the potential of games as assessments of 21st century skills can be realized, there are a number of practical challenges to be addressed. New interactive digital experiences such as games elevate the importance of micro-patterns in data which often reflect variation in strategy or evolving psychological states. Highly granular data about minute human-computer interactions are now available in vast quantities. While the richness of the data holds promise, standard methods by which to turn these data into inferences about knowledge, skills, and attributes are not well-developed. We must be able to identify evidence of our constructs of interest in large files that log the small details of in-game events and we need models for the accumulation of this evidence into estimates of students’ skill
proficiency. Many of our current models of assessment design and delivery will have to be modified, and some completely discarded and re-imagined, in order to conduct game-based assessments of 21st century skills.

Evidence-centered design

A framework to assist with both the conceptualization and implementation of game-based assessment can be found in Evidence-Centered Design (ECD; Mislevy, Steinberg, & Almond, 2003). ECD assists us in developing the assessment argument (Whitely, 1983). It specifies an evidence model that links the output of a learner/player, in this case the data stream from their interaction with a digital environment, and the skill or attributes we are interested in. The evidence model consists of evidence rules and measurement models. Evidence rules refer to the means by which we select particular elements from a student’s work product and apply scoring rules to obtain observable values. Measurement models provide us with the method of combining these observable values to make inferences about our constructs of interest (Mislevy et al., 2003).

In practice, evidence rules and measurement models are often more complex for digital learning environments like games than for traditional assessments. In a digital learning environment, it is not always clear which aspects of the submitted work products (e.g., causal maps, reports of experiments, configured computer networks) should be attended to or how they should be weighted and aggregated to form an overall estimate of skills, knowledge, or attributes.

Persistence

This paper describes efforts to define an evidence model for the attribute of persistence. Task persistence is defined as continuing with a task despite obstacles or difficulty. In the cognitive literature, persistence is generally classified as an element of executive function and thought to be related to self-regulated attention, cognition, and behavior (Anderson, 2002). Persistence may not seem like a distinctly 21st century skill, given that there was a historical review of the literature on measurement of persistence written in 1939 (Ryans, 1939). However, it is often enumerated in lists and discussions of 21st century skills and attributes (Fadel, 2011; Pellegrino & Hilton, 2012), because jobs in the 21st century are increasingly complex, requiring sustained application of effort to complete multifaceted tasks (Andersson & Bergman, 2011).

Persistence is of particular interest and importance because persistence at a young age has been shown to be predictive of many academic and employment outcomes, including adult educational attainment, income, and occupational level (Andersson & Bergman, 2011). The relationship between persistence and academic achievement has been repeatedly documented (Boe, May, & Boruch, 2002; Deater-Deckard, Petrill, Thompson, & DeThorne, 2005; McClelland, Acock, Piccinin, Rhea, & Stallings, 2012).

It is not just persistence with academic tasks that predicts persistence with other academic tasks. Two-year-olds who spent more time trying to open a plexiglass box containing a toy were found to have fewer behavior problems and were more likely to complete school work at age 5 (Sigman et al., 1987). McClelland et al. (2012) report that parents’ ratings of their children’s persistence with difficult toys predicted college completion by age 25, suggesting valid, reliable measures of persistence may help us monitor and intervene with an aspect of learners that can significantly impact their future success.

In order to determine the indicators of persistence to look for in game play, we can examine how other researchers have operationalized it. When directly observing behavior, some have counted the number of times a task is attempted (Foll, Rascole, & Higgins, 2006) while others measure the amount of time during which a child exhibits task-directed behavior (Sigman et al., 1987). Importantly for this paper, Shute and Ventura (2013) defined four indicators to measure persistence in a digital game: time spent on a solved problem (by difficulty), time spent on an unsolved problem (by difficulty), number of restarts for solved problems (by difficulty), and number of re-starts for unsolved problems (by difficulty). It appears the commonalities across studies are time spent on difficult tasks, task completion and attempts after failure.
Purpose and research question

This paper describes research that attempted to determine whether a common factor could be identified under game actions thought to indicate persistence, as revealed by data captured in log files from Poptropica®, a popular commercial game for children. It describes the scoring model and the evaluation of a measurement model that result in a measure of persistence.

Method

The game

Poptropica® is a virtual world in which players explore “islands” with various themes and overarching quests that players can choose to fulfill. Players choose which islands to visit and navigate the world by walking, jumping and flipping. The quests generally involve 25 or more steps, such as collecting and using items, usually completed in a particular order. For example, in Vampire’s Curse Island (see Figure 1), the player must rescue her friend Katya who has been kidnapped by Count Bram. In order to do this, she must navigate to the Count’s castle (eluding wolves on the way), discover the potion for defeating the vampire, identify and mix the potion ingredients, hit the vampire with a potion-tipped arrow from a crossbow, and then find Katya. Apart from the quests, players can talk to other players in highly scripted chats, play arcade-style games head-to-head and spend time creating and modifying their avatar.

![Figure 1. Screenshot from Poptropica’s vampire’s curse island](image1.png)

![Figure 2. Example of log file](image2.png)
The backend of the game captures time-stamped event data for each player. Events include, for example, the completion of quest-related steps, modifying avatars, collecting objects, and entering new locations (see Figure 2). On an average day 350,000 players generate 80 million event lines. Players can play the game anonymously and for free or, with purchase of a membership, and gain access to premium features such as early access to new quests.

Persistence in general requires the presence of a difficult task (Shute & Ventura, 2013), given that easy tasks will not provide the challenge and failure that will allow for the exhibition of persistence. One way to measure difficulty is the percent of players who successfully complete a task. Across islands on Poptropica, completion rates range from 3% to 9%, indicating that these are not easy tasks to complete.

Participants

The final sample consisted of 892 players, 51.2% females and 48.8% males, which is consistent with the near-even gender divide in the overall Poptropica player population. Players ranged in age from 6 years old to 14 years old, with an approximately normal distribution across ages, peaking at age 10 (18.6% of the sample).

The final sample was selected from a larger initial sample. First, a group of players who had at least 500 game events in one of two days was selected. Given that persistence is defined in terms of work towards a particular goal (Meier & Albrecht, 2003), we quickly determined that in order to investigate persistence towards a goal we had to first identify people who were working toward that goal; it would be inaccurate to describe someone as lacking persistence when in fact they were being persistent toward another goal. In this research, we defined persistence as persistence toward completing quests in the game. However, players of the game have varying goals, including social interaction and creating avatars. Therefore, a sample of players with the goal of completing a quest was created by selecting players who completed at least the first purposive action required to complete the quest.

A final examination of distributions by island resulted in the removal of three islands from the dataset due to their low numbers of events required to complete a quest, resulting in outlying patterns of players on quest event-related indicators. Finally, all individuals who had three islands that they made “serious” attempts at completing were combined to form the final sample.

Selection of indicators

The creation of indicators from the log file data was based on the salient elements of measures of persistence in previous research, summarized above, and the data available from Poptropica. Previous research indicated that measurement of persistence required tasks that would be difficult for those attempting them, in order to observe responses to this difficulty. In Poptropica, fewer than 10% of those beginning quests go on to complete them and the number of YouTube walkthroughs and hint pages published by players suggest that the quests are challenging to solve. In order to measure the response to this difficulty, a number of potential indicators were considered in line with the time spent, tasks completed, and attempts after failure. Specifically four potential indicators from each quest were investigated: time spent on quest events, number of quest events completed, maximum time spent on an individual quest event successfully completed, and time spent on the last event prior to quitting the island (unsuccessful quest event). Data was not captured in the game in a way that would allow for a count of number of attempts after failure; this could only be inferred from the time from the completion of one event to the completion of the next event. The variation and relationships among these four indicators were investigated with a small pilot sample. Unfortunately, the data from individual events, both successful and unsuccessful, proved to be highly variable and unreliable as indicators. Therefore, we settled on two indicators per quest: the total time spent on quest-related events and the number of quest events completed. These indicators were computed from the log files for each of the three quests each player attempted to “seriously” complete.
Results

Data cleaning

Examination of the distributions of the indicator variables indicated that they were highly skewed. Given the normality assumptions of the procedures in the study, various transformations of the variables were explored and it was determined that a logarithmic transformation of the number of quest events and a square root transformation of the total time variable resulted in more Gaussian distributions (see Figure 3). The variables were then standardized by island into z-scores in order to account for between-island differences in number of events and time required to complete them.

![Figure 3. Logarithmic transformation of Quest Events indicator](image)

Confirmatory factor analytic model

A confirmatory factor analytic model was run in order to evaluate whether a common factor explained variance across quest indicators. In confirmatory factor analysis (CFA), we specify a model of how the variables are related a priori, and then examine whether the data match what the model would predict. In the case of this research, if these measured variables have a pattern of variance and covariance that is in accordance with what our specified model would predict, that would indicate there is a common construct underlying this group of indicators, and we can then discuss what that construct is and how we can create scores for it.

The first step in a confirmatory factor analysis is the examination of the pattern of correlations. If our indicators all measure a single underlying construct, we would expect moderate to large correlations between them. Table 1 presents the correlations between the indicators for each of the three islands. Correlations are estimates of linear relationships. The raw, non-normal variables do not have a linear relationship. Therefore, the correlations of the transformed variables provide a better estimate of the correlations between the underlying constructs than the raw data.

<table>
<thead>
<tr>
<th></th>
<th>I1 Quest Events (Log)</th>
<th>I1 Total Time (SQRT)</th>
<th>I2 Quest Events (Log)</th>
<th>I2 Total Time (SQRT)</th>
<th>I3 Quest Events (Log)</th>
<th>I3 Total Time (SQRT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I1 Quest Events (Log)</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I1 Total Time (SQRT)</td>
<td>.64</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I2 Quest Events (Log)</td>
<td>.48</td>
<td>.47</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I2 Total Time (SQRT)</td>
<td>.49</td>
<td>.64</td>
<td>.64</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I3 Quest Events (Log)</td>
<td>.42</td>
<td>.39</td>
<td>.54</td>
<td>.55</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>I3 Total Time (SQRT)</td>
<td>.38</td>
<td>.48</td>
<td>.56</td>
<td>.69</td>
<td>.80</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>I1 Quest Events (sec.)</th>
<th>I1 Total Time (sec.)</th>
<th>I2 Quest Events (sec.)</th>
<th>I2 Total Time (sec.)</th>
<th>I3 Quest Events (sec.)</th>
<th>I3 Total Time (sec.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median</td>
<td>15</td>
<td>688</td>
<td>11</td>
<td>397</td>
<td>8</td>
<td>215</td>
</tr>
<tr>
<td>IQR</td>
<td>12</td>
<td>759</td>
<td>10</td>
<td>505</td>
<td>7</td>
<td>305</td>
</tr>
</tbody>
</table>

*Note. I1=Island 1, I2=Island 2, I3=Island 3, SQRT = Square Root, IQR = Inter-quartile range*
A confirmatory factor analytic model was specified with the two indicators of persistence from each of the three islands loading onto a single factor. The error terms for the pairs of indicators from the same island were each correlated because these measures are not independent (completing more events requires more time). Given this hypothesized model, we now want to evaluate it to determine whether it is likely a true representation. In order to do this, we compare the variance-covariance patterns the model would predict to the variance-covariance patterns in the actual data. To make this comparison, we use fit indices that estimate how well the data fits the model. However, because these indices can be sensitive to things like sample size and the size of the model, there are a number of different indices that need to be examined to gain a complete picture of adequacy of fit.

The model fit statistics for this model are displayed in Table 2. The chi-square is significant, which is not desirable, but chi-square is notoriously sensitive to sample size (Kline, 2005), which led to the consideration of alternative fit indices. One way to use the chi-square is to divide it by the degrees of freedom, with values less than 3 indicating good fit (Schermelleh-Engel, Moosbrugger, & Müller, 2003). The Comparative Fit Index (CFI) is an estimate of the improvement of the model over a null model, with a correction for sample size. The Tucker-Lewis Index (TLI; also called the Non-Normed Fit Index) is also an incremental fit index, but with a correction for model complexity. In both cases, values over .95 are indicative of good fit (Schermelleh-Engel et al., 2003; Sivo, Fan, Witta, & Willse, 2006). The Root Mean Square Error of Approximation (RMSEA) is an absolute measure of approximate fit in the population and cutoffs of .01, .05, and .08 have been suggested for excellent, good, and mediocre fit, respectively (MacCallum, Browne, & Sugawara, 1996). The Standardized Root Mean Square Residual (SRMR) is a standardized summary of the average differences between the observed and model-implied covariances. In this case, values less than .10 are indicative of good fit (Schermelleh-Engel et al., 2003).

Examination of the fit of the original model indicated that only the CFI and SRMR met criteria for good fit. Examination of the modification indices suggested that the three indicators of the number of quest events should be correlated. This appeared to be a reasonable modification, given that these three measures could be considered the same measure given across three time periods. Reddy (1992) cautions that ignoring correlated measurement error may lead to inaccurate estimates of coefficients. Correlating these errors resulted in a model with the fit in the second row of Table 2, which indicates good to excellent fit across all indices. This model is displayed in Figure 4.

All of the unstandardized coefficients in the model are significant. The standardized coefficients for the error terms for each indicator provide the percent of variance for each indicator unexplained by the model. For most of the indicators, the persistence variable by itself still leaves a portion of variance unexplained; however, the six indicators do share a significant portion of variance. Examination of the loadings indicates that the time variables appear to have slightly stronger loadings than the completion variables. In addition, the correlation between the two indicators for each island is moderate and significant, indicating that the time taken and events completed on a given island are related above and beyond the variance in persistence. This correlation might indicate the effects of the context of the island, such as whether the island theme was of interest to the player.

Figure 4. Standardized solution for measurement model of persistence
Table 2. Model fit summary

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2$</th>
<th>df</th>
<th>$\chi^2$/df</th>
<th>CFI</th>
<th>TLI</th>
<th>RMSEA</th>
<th>SRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original Model</td>
<td>88.09</td>
<td>6</td>
<td>14.68</td>
<td>.97</td>
<td>.93</td>
<td>.13</td>
<td>.03</td>
</tr>
<tr>
<td>Modified Model</td>
<td>5.16</td>
<td>3</td>
<td>1.72</td>
<td>.99</td>
<td>.99</td>
<td>.03</td>
<td>.01</td>
</tr>
<tr>
<td>Cross-Validation – configural invariance</td>
<td>7.75</td>
<td>5</td>
<td>1.55</td>
<td>.99</td>
<td>.99</td>
<td>.02</td>
<td>.01</td>
</tr>
<tr>
<td>Cross-Validation – metric invariance</td>
<td>15.86</td>
<td>11</td>
<td>1.44</td>
<td>.99</td>
<td>.99</td>
<td>.03</td>
<td>.02</td>
</tr>
</tbody>
</table>

Note. CFI = Bentler Comparative Fit Index; TLI = Tucker Lewis Index; SRMR = Squared Root Mean Square Residual

Although the fit of the re-specified model is very good, the post-hoc modifications run the risk of capitalizing on chance and sampling error in the initial sample. For this reason, an independent sample of 240 players was randomly selected from a different day of play to cross-validate the results. A factorial invariance approach was used, first assessing configural invariance, or the notion that there is the same number of factors (one in this case) and the same free and fixed parameters across both groups. As the fit of this model, displayed in Table 2, indicates, configural invariance was supported, $\chi^2(5) = 7.75$, $p = .26$ suggesting that the error parameters correlated in the model modification step also result in good fit in the independent sample. Next, metric invariance was tested by fixing the factor loadings between the two samples to be equal. Again, invariance was supported, $\chi^2(11) = 15.86$, $p = .15$. The change from the configural model to the metric model, $\Delta\chi^2(5) = 8.12$, $p = .15$, supports the equivalence of the factor loadings across the two samples.

The model can also be used to create factor score estimates for each individual. A number of different methods exist for this computation. In this case, Thurstone’s regression method was employed, in which the observed standardized values are multiplied by regression coefficients (which are obtained by multiplying the inverse of the observed correlation matrix by the matrix of factor loadings. This results in a standardized metric, much like a z-score, with a mean around 0, but where the standard deviation depends on the application. In this case, the overall factor scores have a mean of 0 and a standard deviation of .36.

Relationship to grade

These individual factor scores can be combined to examine group trends as well. Although played anonymously, Poptropica players enter their grade prior to playing the game. Persistence has been shown to increase over the ages spanned by these grades (Hartshorne, May, & Maller, 1929; Lufi & Cohen, 1987), likely as a typical developmental change. Therefore, we computed factor scores for all players in the original sample and examined their relationship to age. As shown in Figure 5, there is a general upward trend in scores across grades, rising from first grade ($M = -.13, SD = .37$) to ninth grade ($M = .13, SD = .34$), for a change of .72 deviations across the nine grades.
Reliability

The internal consistency reliability of the scale, as measured by alpha, was .87. There are essentially only six indicators in this measure so a clear way to increase the estimate here would be to add observations from more islands. However, this would also reduce the number of players for whom we could create estimates because fewer players make serious attempts at four or more islands. The reliability when leaving each indicator out in turn is presented in Table 3.

<table>
<thead>
<tr>
<th>Table 3. Alpha when each indicator removed</th>
<th>Alpha when removed</th>
</tr>
</thead>
<tbody>
<tr>
<td>I1 Quest Events (Log)</td>
<td>.85</td>
</tr>
<tr>
<td>I1 Total Time (SQRT)</td>
<td>.86</td>
</tr>
<tr>
<td>I2 Quest Events (Log)</td>
<td>.85</td>
</tr>
<tr>
<td>I2 Total Time (SQRT)</td>
<td>.83</td>
</tr>
<tr>
<td>I3 Quest Events (Log)</td>
<td>.85</td>
</tr>
<tr>
<td>I3 Total Time (SQRT)</td>
<td>.84</td>
</tr>
</tbody>
</table>

Note. I1=Island 1, I2=Island 2, I3=Island 3, SQRT = Square Root

Discussion

The purpose of this research was to demonstrate the use of data from game play log files to assess a 21st century attribute. Specifically, persistence, or continuing in the face of difficulty, was assessed with a combination of completion and time indicators. Examination of the existing literature on persistence indicated that time spent on difficult tasks, task completion, and continuing after failure were the most common measures of persistence. Similar indicators were created from the log files of the Poptropica game and tested via confirmatory factor analysis. The results suggest there is likely an underlying factor that explains variance across these indicators, and cross-validation on an independent sample confirms the structure and factor loadings. Finally, scores created from these indicators increase across grade levels as we would expect, providing preliminary evidence of validity based on relationships to external variables.

The model fit indicated here reveals that the pattern of variances and covariances in the data very closely fit the model specified. This suggests it is reasonable to discuss a construct that underlies the observed measures. However, in confirmatory factor analysis, one must be cautious in interpretation due to the presence of equivalent models (Kline, 2005). In any confirmatory factor analysis, there are a number of models that are statistically equivalent, that is, will produce the same fit because the models themselves imply the same variance-covariance relationships. It is up to the researcher to provide justification for selecting among these models. In this research, an equivalent model would be to specify a hierarchical model in which the two indicators from each island for their own island-specific measures of persistence that then are combined into a higher-order persistence measure. This model would produce identical statistical results to the model here, and might be chosen if we were interested in a particular research question around comparing islands. However, since we are not, this hierarchical model simply adds more complexity without contributing explanation, so the more parsimonious model was chosen. As Kline notes, “if a single-factor model cannot be rejected, then there is little point in evaluating more complex ones” (2005, p. 211).

A key in the creation of measures is to determine whether they adequately represent the breadth of the construct of interest. Construct underrepresentation occurs when tasks in an assessment fail to include important dimensions of the construct. As described in the literature, persistence is a unidimensional construct and is often measured by a single indicator in research. In this case, we measured both time spent and completion of difficult tasks. As persistence is defined as continuing in the face of difficulty, this appears to cover the breadth of the construct. However, there is also still the clear need for further validation of the measure. Evidence from relationships to external sources, such as rating scales or measures in other settings, are needed to establish that results obtained via this game method correspond to results obtained by other means, providing further support that this is in fact measuring persistence. This lack of validation from external sources limits the conclusions that should be drawn from the measure at this time. However, there are few examples of combinations of indicators from log files of games to measure latent constructs, so this research can be taken as an example of the preliminary work that needs to be done to complete this task.
One important element to note in this analysis is the identification of a sample of players all working toward the goal of completing a quest. Efforts to model the data on the overall player population indicated a general lack of correlation among our indicators of interest. It was the realization that persistence is progress toward a particular goal and then the measurement of just those with a particular goal that resulted in a model with good fit to the data. In fact, the identification of a desired outcome or goal has been identified as a first step in a persistence process (Meier & Albrecht, 2003). This point is important for at least two reasons. First, when researching assessment in games, we must realize that many games are very open-ended spaces and players may be pursuing a variety of goals in game play, including, for example, achievement in the game, exploration, and socializing with others (Bartle, 1996). If any type of inference is to be drawn based on players’ behavior, it is important that their motivations and goals in the game be clear. Second, many people are interested in persistence toward larger goals outside games, particularly completing secondary and post-secondary schooling (e.g., Brown et al., 2008; Kuh, Cruce, Shoup, Kinzie, & Gonyea, 2008). It is important to remember that not everyone enrolling in post-secondary education has the same goal (Provasnik & Planty, 2008). Any research on persistence should take the issue of goals into account.

**Games for assessment, learning, and fun**

This research was completed using a commercial game that was not designed for assessment. This is useful as an exemplar that demonstrates the potential of all kinds of games for measurement and the ability to use data captured from every day interactions to make inferences about players. However, it also led to limitations based on the types of data that were available. Given that the data to be collected from the game were defined long before this research began, events that might have improved this measurement were not tracked. For example, it would have been preferred to have a record of each attempt and failure individuals made at an event. When the player is trying to get past a wolf, it would have been nice to have a count of how many times the wolf chomped the player and how many times the player tried again to get past. This data was not available, only the eventual success at getting past the wolf is recorded along with the time spent trying to do so. This time can serve as a proxy for the number of tries (more tries is likely related to more time) but is not a direct measure. Time can also be subject to things like distraction, daydreaming, and mental processing speed.

This lack of identification of appropriate evidence for our constructs is one danger of not addressing key assessment constraints from the beginning of game design (Mislevy, Behrens, DiCerbo, Frezzo, & West, 2012). If we are interested in making inferences about players, a better approach is to bring together game designers and assessment experts from the beginning of the design process. Early design sketches should include not just game features, but preliminary assessment arguments. Essentially this partnering from the beginning allows for the building of better links between what players do, what elements of their work product are collected, how they are scored, and how we accumulate the information to make inferences.

Researchers also need to consider how learning and assessment interact in games. Discussions of gaming in education often focus on the evidence (or lack thereof) that games can enhance learning. This paper has not addressed that question. Rather, it has demonstrated the use of games to gather information about a student attribute that would be difficult to assess with traditional assessment. However, a next step for this research might be to investigate whether persistence changes over the course of continued game play.

Future research, apart from the needed external validation of the measure described here, might then focus on how games might be designed to foster and/or train persistence. It might be hypothesized that elements of game play reward players for persisting in their efforts, thereby increasing the likelihood that they will persist more in future interactions (Ventura, Shute, & Zhao, 2012). Although its definitions as a personality trait suggests for some that persistence might be immutable, research has shown that it is influenced by a variety of task and personal factors (Kamins & Dweck, 1999; Reiher & Dembo, 1984; Schunk, 1983; Thomas & Pashley, 1982). It appears that feedback that encourages students to attribute success to hard work as opposed to internal factors, has shown to be effective in increasing task persistence in other environments. Given games already are efficient at providing feedback to players, designing them to give this specific type of feedback would be an interesting challenge. Finally, the relationship between persistence and factors such as prior content knowledge and prior experience in a given learning environment influence persistence.
Games and learning analytics

While games have many unique qualities, they represent just one digital environment that can collect data. Today’s learning management systems allow us to collect data from a wide swath of sources. This can include usage information about logins to learning management systems, posts to discussion boards, interactions in simulations, along with traditional assessments. All of this ubiquitous data, collected in a largely unobtrusive manner, allows us to make inferences about students in new ways (DiCerbo & Behrens, 2012). The emerging field of learning analytics focuses on gathering, analyzing, and applying all of these new data types to the issue of understanding and improving learning (Siemens & Gasevic, 2012).

Learning analytics includes the mapping of knowledge domains and then the evaluation of learner activity in relation to those maps. Assessment becomes a continual activity, as new data is collected, it is used to update a students’ profile of knowledge, skills, and abilities. Content and intervention, when needed, can be recommended and delivered in response to this changing profile, rather than in lock-step as it is currently (Siemens & Long, 2011). Games then, may provide just one piece of evidence that can provide information into such a profile. In the case presented here, a measure of persistence from a game like Poptropica might be combined with a persistence measure from viewing course videos and one from a homework system. A measurement of low persistent might trigger an alert to an instructor, and might also inform an instructional system to appropriately adjust feedback given to the student.

Concluding thoughts

The promise of game-based assessment is that we can use data from in-game actions to make inferences about players’ knowledge, skills, and attributes, allowing us to make use of information in the ocean of data produced by daily digital interactions with software (DiCerbo & Behrens, 2012). In order to do this, we need to be able to identify evidence from the log files produced by these interactions and accumulate that evidence to produce valid, reliable measures. This paper demonstrates one promising effort to create a model of persistence from young players’ actions in Poptropica using a combination of data mining and factor analytic techniques. While the loadings and indicators presented here are specific to Poptropica, the ideas could be applied across settings. Indicators of time spent on and progress through difficult tasks can be gathered from a variety of games and other digital activities. These indicators can be standardized and combined through factor scores or other methods of aggregation. To summarize, for others interested in assessing persistence, this work suggests, they should (1) clearly identify the goal individuals are working toward, (2) measure how much progress was made toward the goal, and (3) measure how long players spent trying to complete difficult tasks. These measures or indicators can be combined with relatively common statistical techniques, including factor analysis. The continued exploration and combination of new and existing methods to extract information from data streams may fundamentally change how we think about assessment.

Acknowledgments

The author wishes to thank Dan Murphy, David Shin, Steve Ferrara, and the anonymous reviewers for their suggestions which greatly improved this work.

References


Undertaking an Ecological Approach to Advance Game-Based Learning: A Case Study

Mamta Shah* and Aroutis Foster
Drexel University, School of Education 3141 Chestnut Street, Philadelphia PA 19104 USA // mamta.p.shah@drexel.edu // aroutis@drexel.edu

*Corresponding author

ABSTRACT
Systematic incorporation of digital games in schools is largely unexplored. This case study explored the ecological conditions necessary for implementing a game-based learning course by examining the interaction between three domains (the innovator, the innovation, and the context). From January-April 2012, one in-service teacher learned and applied the Play Curricular activity Reflection Discussion (PCaRD) pedagogical model using the game RollerCoaster Tycoon 3 in a 5th and a 6th grade classroom to support student learning of systems thinking. Data sources included a knowledge test, interviews, and participant and video observations. The teacher (innovator) successfully adopted the PCaRD model (innovation) as a pedagogical guide for teaching systems thinking with the game in the school (context). Students had statistically significant gains in introductory systems thinking knowledge. Conclusions and implications are discussed for integrating games using our ecological approach.

Keywords
Game-based learning, Game integration, PCaRD, Ecological approach, Case study

Introduction
There is growing willingness among teachers and administrators to use digital games in schools (Millstone, 2012). However, despite the rising number of empirical studies that report an improvement in student academic achievement and motivation as a result of using games (Barab, Pettyjohn, Gresalfi, Volk, & Solomou, 2012; Kebritchi, Hirumi, & Bai, 2010), the evidence for it in many academic domains remains slim (Young et al., 2012). It has been suggested that research studies need to consider the dynamics of teacher intervention, social contexts, classroom environments, learning activities, and the alignment of the games used to facilitate an adequate understanding of the educational value of games (Young et al., 2012). This article reports results from a case study utilizing the Play Curricular activity Reflection Discussion (PCaRD) pedagogical model (Foster, 2012) in a K-8 private school over three months. We explored the ecological conditions of how a game-based learning course can be incorporated in classrooms through partnership with teachers.

First, a literature review on game-based learning is presented. This section also describes the rationale for undertaking an ecological approach to integrate digital games in schools and the theoretical framework adopted for the current study. Second, the research question and methodology are described. Third, the results section illustrates our ecological approach, highlighting the interaction between the participating teacher as the innovator, PCaRD as the game-based learning innovation, and the participating school as the larger context. Fourth, the concluding sections discuss the benefits of the ecological approach with PCaRD for the integration of games into K-12 classrooms.

Literature overview
Trends in game-based learning
Current trends suggest that many classroom-based game studies examine the effects of games on student achievement and motivation (Kebritchi et al., 2010; Papastergiou, 2009; Tobias & Fletcher, 2012). For instance, Kebritchi and colleagues (2010) examined high school students’ mathematics achievement after using the game Dimension M and found gains from pretest to posttest. Papastergiou (2009) assessed and found significant effectiveness of a computer game called LearnMem1 on high school students’ knowledge of and motivation to learn computer memory concepts.
Conversely, game studies conducted afterschool focus their attention on interpreting the process of student learning through games (Foster, 2011; Squire, DeVane, & Durga, 2008). For instance, Foster (2011) used RCT3 and showed middle school students’ development of motivational valuing and disciplinary knowledge construction for microeconomics principles associated with operating a business. Squire and colleagues (2008) used Civilization III and showed how learning in games motivated disengaged high school students, leading to an interest in academic practices and development of world-history knowledge. Studies in both settings indicated the appeal of digital game-based learning for improving student learning and motivation; however, neither the in-school or afterschool studies explored the role of classroom dynamics or teachers in these learning environments.

Some recent studies have indicated that teachers have the potential to augment the effect of games on students’ interdisciplinary knowledge construction and motivation to learn (Barab et al., 2012; Silseth, 2012). For instance, Silseth (2012) examined how the interactional episodes of one high school student with his peers and a teacher in a gaming context contributed to his understanding of the multifaceted nature of the Israeli-Palestinian conflict. Barab and colleagues (2012) described the role of one teacher facilitating two curricula (game-based versus story-based) on persuasive writing and its impact on student engagement with the topic. In both studies, teachers were knowledgeable about the game and used it as a resource to support students. Teachers performed several roles, which included (a) being an expert guide to help students navigate the nuances of the game and make connections with the learning objectives, (b) adopting multiple pedagogical approaches (e.g., instruction, discussion, observation) to invoke student reflection and provide feedback, and (c) aiding students to understand the relevance of their knowledge beyond the course.

Teacher intervention can be crucial in facilitating game-based learning; however, there is a lack of knowledge available in relation to the pedagogical processes involved in implementing games in classrooms (Hanghøj & Brund, 2011). Watson and colleagues (2011) showed that even when a teacher is knowledgeable about a game, a lack of systematic guidelines on how to implement games in classrooms results in game incorporation through trial and error, and assessments that do not fully capitalize on student engagement with the topic and the strengths of the game as a curriculum. Researchers argue that teachers need more holistic support if games are to be effectively integrated in academic courses in schools (Jaipal & Figg, 2009; Watson et al., 2011).

**Game Integration in Schools**

Extant literature has produced a list of factors affecting the integration of games in schools: (a) working around the daily school schedule and teachers’ pressures of content coverage (Baek, 2008), (b) ensuring that the school has a suitable physical and technological infrastructure (Ritzhaupt, Gunter, & Jones, 2010), (c) seeking training opportunities and administrative support for teachers (Kenny & Gunter, 2011), (d) identifying and adopting games that align with the curricular goals (Hirumi, 2010), (e) guiding teachers with appropriate game-based learning pedagogical models (Gros, 2010), and (f) supporting students with different levels of readiness to adjust to the use of games for learning in the school context (Rice, 2007). According to Ketelhut and Schiffter (2011), there is a dearth of studies that jointly examine the effectiveness of employing digital games and its interaction with school conditions and implementers. The authors argue that these gaps should be addressed in order to facilitate significant digital game integration in schools.

One approach that considers some of the gaps identified is the Game-Network Analysis (GaNA) framework. GaNA was developed for enabling teachers in introducing game-based learning in classrooms by guiding them in game analysis and game integration within a curriculum (Foster, 2012). GaNA is a combination of the Technological Pedagogical Content Knowledge (TPACK) framework (Mishra & Koehler, 2006) and the Play Curricular activity Reflection Discussion (PCaRD) model (Foster, 2012). TPACK provides a lens for game selection and analysis; that is, it helps teachers approach the game as a curriculum with constraints and affordances for technology, pedagogy, and content (Foster, Mishra, & Koehler, 2012). PCaRD is a pedagogical model that aids teachers in incorporating games in classrooms to achieve curricular goals and facilitate students’ engagement in academic domains (Foster & Shah, 2012). PCaRD includes the Inquiry, Communication, Construction, and Expression (ICCE) framework to foster transformative learning anchored in the game (Dewey, 1902; Shah & Foster, in press). Thus, GaNA provides the frame a teacher needs within their ecological context to focus on the pedagogy and content of games as well as the process to use and apply games in classrooms. In this article, we focus on PCaRD as a pedagogical innovation used by a teacher to teach a game-based learning course in a school context.
PCaRD provides a systematic pedagogical process that is flexible enough to accommodate and guide teachers in using games for teaching. Researchers (Foster, 2012; Foster & Shah, 2012) conducted a study supporting teachers and implementing the PCaRD model to facilitate students learning in mathematics and science. Results indicated significant student knowledge construction and the development of interest in the content. The teachers initially had little confidence, but commitment to use games. They became more confident and comfortable integrating games with PCaRD in a school with limited technological infrastructure and support. PCaRD served as part of an ecological approach for establishing the conditions for integrating games into classrooms.

**Ecological approach**

Scholars argue that technology adoption should be approached from an ecological perspective in which the relationships between multiple factors within the larger school context are considered (Nardi & O'Day, 1999; Zhao & Frank, 2003). For game-based learning, Klopfer (2011) has argued for an ecological approach to aid the practice of using games in a comprehensive manner, with consideration for realistic learning environments. An ecological approach is essential for adding relevance to claims made about the educational use of games in context. Thus, the authors adopted Zhao and colleagues’ (2002) ecological technology integration model.

Zhao and colleagues’ (2002) model consisted of 11 interacting factors grouped under three domains: the innovator, the innovation, and the context that impact the degree to which technology projects are integrated well in classrooms. Thus, in this paper, the conditions for using games in classrooms were established as follows: innovator—teacher in the classroom; innovation—PCaRD for game-based learning; and context—the school in the study (See Figure 1). We used all the 11 factors (Zhao et al., 2002) in considering the interaction between the innovator (teacher’s knowledge of technology, technology-teacher pedagogical belief compatibility, teacher-school dynamics negotiability), the innovation (its distance from school culture, existing practice, available technological resources, and its dependence on human and technological resources) and the context (human and technological infrastructure, and social support within the school).

This paper aims to emphasize that successful adoption of game-based learning requires an ecological approach. Such an approach is missing from the current studies, causing hindrances to teachers’ and researchers efforts at supporting student learning with games. The case study presented in this paper addresses these gaps in the field by (a) offering one example of the domains to be considered in undertaking an ecological approach, (b) ascertaining the effectiveness of our approach on student learning. Thus, the following research question was investigated, “What conditions are necessary in a satisfactory implementation of a game-based learning course for teaching systems thinking in a K-8 school?”

![Figure 1. Theoretical framework for determining conditions necessary for using digital games in K-12 schools](image-url)
Methods

An instrumental case study (Stake, 1995) was undertaken to examine the conditions necessary for game-based learning using an innovation – PCaRD, in partnership with an innovator - in-service teacher in a K-8 school - context.

Context and participants

In the spring of 2011, one female science and technology teacher with less than five years of K-12 teaching experience expressed interest in implementing PCaRD in her courses at a K-8 private school in a US Northeastern suburban city. As a Director of Design and Innovation at the school, the teacher’s goal was to develop and test programs focusing on higher-order thinking within a playful environment. Between the summer and fall of 2011, the teacher and the researchers created an introductory systems thinking course to support her students. In the winter of 2012 (January-April 2012), the teacher offered the course in a computer lab without change to eleven grade 6 students and ten grade 5 students for a total of 21 middle school students in two classes. All students (8 females, 13 males) completed the course. They identified as Caucasian American. These students averaged less than 2-hours of gameplay as compared to the US national average of 7-hours each week for children between 8-18 years old (Rideout et al., 2010).

Course description and goals

The goal of this course entitled ‘Thinking in Systems: a playful approach’ was to use games to teach systems thinking with the game as an example of a system. A system was defined as a set of interacting elements (objects, properties, behaviors, and relationships) that form an integrated whole with a common purpose. Whereas the game provided a platform to demonstrate system dynamics, the PCaRD methodology provided a structure for the weekly lessons for emphasizing specific aspects within systems. Additionally, course activities designed through PCaRD involved ICCE opportunities that allowed students (a) to examine how the elements within the chosen games fit together to form playable systems and (b) to learn how to apply the understanding of basic system principles to control the quality of interactions within systems. Examples of essential questions that guided the design of course activities were, “What is a system and how is a game a system?” and “What is the meaning of control and feedback in systems and games?”

The course was developed to facilitate students in developing the knowledge and skills in analyzing a system. Examples of knowledge types included, foundational-knowledge (e.g., feedback sources for assessing system efficiency), meta-knowledge (e.g., properties and behaviors of system parts to foster problem solving), and humanistic-knowledge (e.g., empathy-related decisions to manage a system). Example of skills included the ability to identify elements of a system, to understand the existence of several sub-systems, and to identify feedback effects in a system.

Data sources

Quantitative data sources included a 9-item pre-post knowledge test for systems thinking which was developed in partnership with the teacher based on the course goals. To aid in item construction, the researchers played RCT3 before the start of the study using the TPACK framework (Foster, 2012). The test included questions based on foundational-knowledge (e.g., identifying feedback sources), meta-knowledge (e.g., understanding the role of feedback), and humanistic-knowledge (e.g. negotiating feedback) about systems thinking in the context of the amusement park in RCT3 (See Table 1).

<table>
<thead>
<tr>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>What sources will tell you if your amusement park is making profit, efficient and successful?</td>
</tr>
<tr>
<td>Suppose new guests were not coming to your park, what could be the possible reasons for this problem?</td>
</tr>
<tr>
<td>What are the essential parts or elements that make up a good and a successful park? Identify at least 5 parts.</td>
</tr>
</tbody>
</table>
Qualitative data sources included semi-structured (pre-course, monthly, delayed-post) interviews with the teacher. Some questions included “What is the role of technology in teaching and learning in your view?” “Please describe your teaching philosophy” “How appropriate is RCT3 for learning systems thinking in terms of content?” and “What are your insights about using PCaRD as you are implementing it?” Weekly in-class participant and video observations, and memos from after-class meetings were also maintained to gain insights about the overall development of the course, teachers’ emerging proficiency of integrating PCaRD through RCT3, and the overall fit of the project in the participating school context. Pre- post- and delayed-post interviews were conducted with students to further the researchers’ understanding of students’ knowledge construction, motivational engagement with systems thinking, and their experience in a game-based learning course using PCaRD. Examples of interview questions included, “What did you pay most attention to while building, designing and managing your park? How did that affect the success of your park?” and “What did you like about this game-based class?”

**Game environment**

Students played RCT3, a commercial entertainment CD-ROM single-player simulation strategy game. It is an interdisciplinary game in which players are allowed to experience a career in building and managing an amusement park within numerous scenarios (See Figure 2). The game was selected based on prior work (Foster, 2011, 2012) and its relevance to the content area in this study.

RCT3 exposed players to the systems thinking concepts in the context of managing an amusement park. For instance, players navigate between a birds-eye view and a street-view of their amusement park as they become engaged in different sub-system activities such as building rides, hiring staff, managing finances, attending to the needs of the park visitors, and monitoring overall park activities. Players were required to satisfy predetermined objectives (e.g., attaining a certain park rating) using finite resources (e.g., money, space) in order to progress. Objectives provided an overarching focus to player’s actions and guided them towards discovering the many system variables and their interactions in managing their park (See Figure 2).

![Figure 2. A screenshot from RCT3 indicating scenario objectives, player actions, and feedback](image)

**Procedure**

Over one term, the researchers worked extensively with the teacher, coaching and modeling to apply PCaRD to teach the course. This included supporting the teacher through face-to-face meetings and online-communication in integrating PCaRD as she planned her lessons. Furthermore, a researcher was present throughout the study to offer in-class support. Teacher-researcher meetings were held after each lesson to reflect on the significant occurrences for the day, to discuss concerns with the PCaRD implementation, and to plan for the forthcoming lesson. Lastly, the researchers encouraged the teacher to increase her familiarity with RCT3 (before and during the intervention) through game analysis around technology, pedagogy, and content for systems thinking.
The weekly course was offered for 90 minutes each for 5th and 6th graders. Typically, the PCaRD session began with naturalistic game **playing** (30-40 minutes), allowing player-game-peers-instructor interaction. **Curricular activities** (20-30 minutes) followed, in which the teacher used the game and students’ play experiences as an anchor to develop problem-based activities for achieving course goals. Prompt-based **reflection** (10-15 minutes) followed, in which individual students articulated how they applied their academic and personal knowledge to progress in the game. Closure to the session was brought by a **discussion** (10-15 minutes) in which students reconvened as a large group to express their opinions, seek explanations to their questions, and learn to make connections with the systems thinking concepts. Additionally, the teacher would address overall learning goals based on her observations of what was learned and what needed more attention. Each phase within PCaRD built upon the other to scaffold students’ learning experiences with opportunities for ICCE. Lastly, every PCaRD session was designed based on insights gained from the previous session and the curricular goals of the forthcoming week.

**Data analysis**

Match-paired *t*-tests were used to measure the change in students knowledge about systems thinking. Student interviews were analyzed using grounded theory (Charmaz, 2006) to understand their knowledge construction (foundational-knowledge, meta-knowledge, and humanistic-knowledge) and motivational engagement with systems thinking, and experience with PCaRD. Classroom observations, interviews, researcher memos, and curricular materials obtained from the teacher were also analyzed as part of the case study to document evidence satisfying and indicating the interaction between the three domains: the **innovator** (the participating teacher), the **innovation** (PCaRD), and the **context** (the participating school). In particular, the researchers examined the teachers’ understanding of (a) PCaRD as a pedagogical guide to implement games in classrooms and the school context, plan instruction, and support student learning (b) RCT3 as a curricular tool with its constraints and advantages for teaching systems thinking. This allowed in the triangulation of data to inform the results.

**Results**

We present the results in three sections to illustrate the ecological approach for addressing our research question about the conditions necessary in a satisfactory implementation of a game-based learning course for teaching systems thinking in a K-8 school. For instance, in the first section we describe the factors associated with the innovator for successfully integrating a technological innovation. This is followed by a description of the nature and extent to which the participating teacher satisfied those factors for implementing the game-based learning course. This is repeated for sections devoted to the domains of the innovation (PCaRD) and the context (the school).

**The innovator (The Teacher)**

Successful technology integration projects are closely related to the innovator’s comprehensive knowledge of the technology, adoption of technology that is consistent with her or his pedagogical beliefs, and ability to negotiate the social dynamics within the school (Zhao et al., 2002). As an educational technology, the game acts as a curriculum with inherent affordances and constraints in terms of pedagogy and content (Foster & Mishra, 2011) for facilitating students’ understanding in a given discipline. Therefore, in order to teach meaningfully with games, a teacher needs to acquire an understanding of students’ everyday knowledge including their knowledge of the game, knowledge about the pedagogical practices in the school, content knowledge, and knowledge which is specific to the game (e.g., relevant game dynamics, genre, specialized knowledge embedded in the game). The participating teacher believed that technology is a tool and its role was to transform learning. In an interview, she expressed her beliefs about using technology for instruction:

“…I am not using technology [because] I have an iPad or whatever, [b]ut if I think it will add another dimension. And I really think that [students] should learn a lot by doing things intuitively…than just you know just learning the equation. So, technology is very good [when] it brings more opportunities for intuitive learning.”
In addition, the teacher believed that RCT3 afforded students with hands-on experiences and a personally relevant context for learning about systems thinking. The game met her expectations of engaging students in intuitive learning about systems thinking:

“[Although] it is not obvious for kids to understand the relationship between parts in an amusement park [in RCT3]… I still think it is a very good way to teach systems thinking and to show them that even something like an amusement park is a system.”

RCT3 played a crucial role in accomplishing the course goals because it provided the basis for situating students’ learning of systems thinking. A well-rounded knowledge of RCT3’s potentials and limitations was also important for planning and implementing the course lessons using PCaRD, as the teacher understood that the relationship between parts in the game was not obvious:

“It is very important to [know] how the course goals can be mastered through the game. I had to make sure that all the concepts of system thinking that were part of my curriculum could be demonstrated through [RCT3] and that it provided added value to the curricular activity.”

The teacher’s knowledge of RCT3 evolved over the period of the intervention. As she increased her familiarity with RCT3 using different methods of game analysis (e.g., playing, observing, conversing) (Aarseth, 2003), she progressed from having a technologically functional understanding of the game, to being skeptical about the relevance of RCT3 to support students’ understanding of systems thinking, and finally gaining confidence in proactively identifying the potentials and overcoming the limitations of RCT3 for teaching systems thinking with PCaRD (See Table 2). PCaRD facilitated the connection between RCT3 and the teacher’s goals.

The innovation (The PCaRD model)

Several factors of an innovation influence its acceptance within the school: its distance from school culture, existing practice, available technological resources, and its dependence on human and technological resources (Zhao et al., 2002). In this study, the game-based learning course was implemented using the Play Curricular activity Reflection Discussion model (PCaRD), a technological and pedagogical innovation for aiding teachers in incorporating digital games in classrooms and supporting students in constructing foundational-knowledge, meta-knowledge, and humanistic-knowledge.

<table>
<thead>
<tr>
<th>Phase of intervention/ Frequency of game analysis</th>
<th>Mode and focus of game analysis</th>
<th>Level of game competence acquired/ Resulting outcome(s)</th>
<th>Changing confidence about the game’s relevance for the course</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior to inception- Initial Weeks/ Infrequent</td>
<td>Game playing and observing a child play RCT3 at home</td>
<td>Basic operational/ Installing RCT3 on to the computers and resolving technical issues (e.g. creating special logins)</td>
<td>“RCT3 is an interesting game”</td>
</tr>
<tr>
<td>Initial Weeks- Middle of Intervention/ Occasional</td>
<td>Game playing at home Speaking with researchers Observing and conversing with students as they played Watching YouTube videos of specific game actions (e.g. ride building)</td>
<td>Limited knowledge/ Experiencing difficulties in integrating RCT3 while planning lessons and supporting students’ play.</td>
<td>“RCT3 does not provide the content for systems thinking. [RCT3] does not make it obvious for the students to understand the relationship between parts in an amusement park”</td>
</tr>
<tr>
<td>Middle of Intervention- Final Weeks/ Frequent</td>
<td>Game playing in the school lab Playing with specific curricular goals in mind (e.g. Playing RCT3 before introducing the concept of feedback)</td>
<td>Emergent knowledge/ Becoming aware of RCT3’s potentials and limitations in the context of the course</td>
<td>“RCT3 provides different feedback parameters to the player; however, you can play the game without paying attention to most of...”</td>
</tr>
</tbody>
</table>
Making conscious decisions about how to use RCT3 while planning lessons.

Therefore, in order to explain the meaning of feedback in a system, the curricular activity [must] be designed to draw students’ attention to the feedback mechanisms in RCT3."

The teacher’s initial goal was to experiment with different approaches in game-based learning. Consequently, PCaRD was welcome within the participating school’s culture. Except for purchasing RCT3, the teacher did not have to request additional technological resources to use PCaRD. PCaRD also aligned well with the teacher’s pedagogical approaches. This was reflected in her understanding of the model, which developed organically as she implemented it (See Table 3).

As her comprehension of RCT3 and PCaRD deepened, the teacher became competent at incorporating relevant features of the game in her lesson plans and scaffolding student understanding during class. In one curricular activity, which focused on the topic of feedback, two types of feedback were first introduced: balancing and amplifying. Next, the teacher illustrated the feedback types using examples from students’ personal lives (e.g., sibling interaction) and school experiences (e.g., a game of basketball) – making connections with school content and student experiences using ICCE. Additionally, the teacher invited students to share examples of feedback, which were locally situated in their personal and school life contexts, and from their experience in RCT3. The teacher used a variety of contexts through PCaRD to introduce the concept before students were assigned to document the nature and sources of feedback.

<table>
<thead>
<tr>
<th>Month</th>
<th>Play</th>
<th>Curricular activity</th>
<th>Reflection</th>
<th>Discussion</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>Students engage in intuitive and tacit learning</td>
<td>Students are provided with labels for understanding what they are learning They bridge the intuitive and conscious aspects of learning</td>
<td>Students translate their experience into words</td>
<td>Students expand their knowledge by sharing and learning from others experiences</td>
</tr>
<tr>
<td>February</td>
<td>Students experience flow as they interact with the game in a way where no one interferes</td>
<td>Students are introduced to the lesson goals through an activity. They acquire a specific focus</td>
<td>Students make the connection between what they learnt in the curricular activity and the play using the new terminology</td>
<td>Students become aware of their feelings and learning by reflecting individually. They discuss and listen to opinions of others in order to learn from each other</td>
</tr>
<tr>
<td>March</td>
<td>Students engage with the game from their own natural curiosity and they interact with their peers</td>
<td>Students are engaged in an activity designed for addressing the essential questions for the lesson They gain a new set of tools to think with</td>
<td>Students use the newly acquired toolbox to make personal connections between play and curricular activity</td>
<td>Students broaden their perspective by being exposed to reflections of their classmates. They gain insights for improving their own learning experience. They tie their entire experience</td>
</tr>
</tbody>
</table>

After the completion of the course, the teacher summed her thoughts about PCaRD for supporting teachers and student participation in a game-based learning course in a school setting:
“[T]he use of PCaRD with its systematic approach was very useful when it comes to implementing video games in the classroom. Since this was the first time for my students to play a game in a learning setting – the structure helped them to see how the game is related to their learning, it disciplined them time-wise, and gave them an opportunity to share experiences. The advantages of the PCaRD model when using a game within a lesson is its ability to create structure when developing the lesson plan, helping with keeping the narrative throughout the lesson’s different stages and making the connection between one lesson to the next.”

The Context (The school)

Technology integration projects are affected by the human and technological infrastructure, and social support within the school. For teachers to capitalize on the affordances of game-based learning environments in schools, support needs to be provided across the context from school leaders such as principal, technology coordinators, and peers (Halverson, 2005). Using a game also requires consideration for technical and logistical issues such as the existence of appropriate hardware and software and sufficient access to it for the users.

The school culture facilitated instruction necessary for game-based learning and teacher development. There were few disruptions to the class schedule and technological infrastructure. There was support for teachers’ in developing expertise in implementing games using PCaRD. There was uninterrupted access to the computer lab and its resources. Students were also observed playing RCT3 in the computer lab during their free time.

According to the teacher, this was beneficial because students would come to the systems thinking course with additional insights for sharing with their peers during play and the discussions. The teacher also began playing the game during her free time. The lab was equipped with eighteen networked desktops and an interactive white board (Figure 3). Although an in-house technology person was not appointed, technical support was promptly available upon the teacher’s request. Few technical issues were experienced and were resolved in the first two weeks of the study. On several occasions, students’ saved games disappeared from the computers unexplainably. However, for most students, this provided an opportunity to start over and create an improved amusement park. Interestingly, the teacher indicated that the school principal had shown a keen interest to continue applying PCaRD in the following academic year with the involvement of additional teachers.

Student Learning

Both 5th and 6th grade students made statistically significant gains in the systems thinking knowledge test (See Tables 4 and 5). They developed a simple understanding of systems as a whole a set of interacting elements, as is visible in their descriptions provided during post-interviews.

Grade 6 students:
“System is how things work and how they make things work. Clocks, computers, our bodies are systems”
“A system is like a group of things that work together and contribute to one thing. In RCT3 you’d have rides, food stands, trees, employees to build an amusement park...and for people to have fun”

Grade 5 students:
“Can’t be one thing working by itself. It has to work like a team.”
“[Things are] connected. For e.g. bike is a system. pedal[ing] makes the wheels move”

Interpretive findings suggest that students found it difficult to see the personal relevance of systems thinking beyond the course. That said, at those grade levels, it was the intention of the teacher to introduce them to the concepts of systems thinking as preparation for courses in higher-grade levels.

Table 4. Descriptive Statistics of Systems Thinking

<table>
<thead>
<tr>
<th>Source</th>
<th>N</th>
<th>PreMean</th>
<th>PreSD</th>
<th>PostMean</th>
<th>PostSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systems Thinking Knowledge</td>
<td>20</td>
<td>14.80</td>
<td>2.60</td>
<td>17.50</td>
<td>2.98</td>
</tr>
</tbody>
</table>

Table 5. Paired t-tests Analysis Of The Knowledge Test For The Overall Sample

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>t</th>
<th>p</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Post Systems Thinking Knowledge</td>
<td>19</td>
<td>4.011**</td>
<td>0.001</td>
<td>0.96*</td>
</tr>
</tbody>
</table>

Note. P < .01**, R² = 0.18*

Upon inquiring students’ experiences with PCaRD, students expressed their appreciation for the interactive nature of the class as opposed to passive instruction and grading based on abstract testing. Students also valued the chance to work on individual and group challenges during the curricular activities and believed that these activities fostered their creativity. Lastly, they shared instances where they used their insights from reflection and discussion sessions to improve their game play.

Discussion

This study investigated the ecological conditions that may be necessary in a satisfactory implementation of a game-based learning course for teaching systems thinking in a K-8 school. There was a good fit between the innovator, the innovation, and the school context- the three domains considered in the ecological approach undertaken (Zhao et al., 2002). The teacher’s strength as an innovator grew from her (a) pedagogical beliefs being consistent with the game used, and (b) her ability to reflect and improvise while implementing PCaRD, a systematic process she valued. The PCaRD model was developed as a systematic, albeit adaptive pedagogical guide to aid teachers interested in adopting game-based learning in schools, and it was characterized by low-moderate reliance on technological and human resources. Lastly, the school offered up-to-date technological infrastructure and encouragement towards offering an innovative course. Undertaking an ecological approach allowed overcoming the gaps identified in game studies (Papastergiou, 2009; Squire et al., 2008); that is, the approach aided in focusing on both the process and outcome of the game-based learning intervention on student learning and simultaneously its interaction with teacher roles and school dynamics.

Effective game integration requires educators to be able to decipher the relationship between a game, the achievement of curricular goals, and its fit within the school context (e.g., physical infrastructure) prior to and during its educational use. Our approach aided the teacher in developing a flexible knowledge of RCT3 in terms of its pedagogical affordances and embedded content for systems thinking in the context of its use (e.g., course goals, existing practices, middle school) (Foster & Mishra, 2011). Additionally, the teacher integrated her game knowledge with the pedagogical structure of the course- PCaRD (Foster & Shah, 2012), which adapted to the teacher and students’ emerging proficiency in the game-based learning course (Baek, 2008). Together, this aided the teacher in augmenting the effect of the game on student learning (Tüzün et al., 2009), and enhancing their motivation to learn about systems thinking by bridging students’ academic, personal, and game experiences (Barab et al., 2012).

When school leaders including administrative and teaching staff gain an appreciation for game-based learning, they can confidently adopt the challenge of innovating the learning environments in the existing school contexts. According to Owston (2007), teacher and student support of the innovation, teacher perceived value of the innovation, teacher professional development, and principal’s approvals are the essential conditions for sustaining a
classroom innovation. In this study, the participating school context was a private school and the principal was excited about the prospects of game-based learning for the students. The participating teacher, as the Director of Design and Innovation had more freedom to learn about and introduce an innovative course, an infrequent autonomy available to teachers (Kenny & Gunter, 2011). The school also supported the teacher with essential technical and logistic support, which is beneficial for integrating games in schools (Ritzhaupt et al., 2010). Although it is important that a successful application of game-based learning requires fulfilling the conditions proposed by Zhao and colleagues (2002), our approach is robust enough to be applied satisfactorily even when one of the conditions are not strongly supported (Foster et al., 2011). A game-based learning pedagogical model like PCaRD when applied holistically will facilitate success with commitment from teachers, administration, and students.

Implications and limitations

There are implications and limitation for our study involving the ecological approach for the use of games in classrooms and for learning. First, students did better in a post-test than they did in the pre-test, but learning is to be expected of any educational intervention. Qualitative evidence suggests student satisfaction was important in supporting their interest and knowledge gain. However, we do not have evidence of differences between this group and previous years where no video games were used for understanding the impact of the innovation. This study was exploratory. Thus, we recommend further research for developing a more conclusive understanding of the impact of this game-based learning ecological approach. Second, teachers may use PCaRD as a guide in the ecological adoption of game-based learning in schools. Since PCaRD is a part of larger framework, GaNA may be used as a guide to enhance a games’ potential as a curriculum, identify possible learning experiences to engage students in a subject area, and integrate the game systematically in academic courses to support students’ development of foundational-knowledge, meta-knowledge, and humanistic-knowledge.

References


An Investigation of the Interrelationships between Motivation, Engagement, and Complex Problem Solving in Game-based Learning

Deniz Eseryel¹*, Victor Law², Dirk Ifenthaler³, Xun Ge¹ and Raymond Miller¹

¹Department of Educational Psychology, University of Oklahoma, USA // ²Program of Organization, Information, and Learning Sciences, University of New Mexico, USA // ³Department of Courses & Quality, Open Universities Australia, Australia // eseryel@ou.edu // vlaw@unm.edu // dirk@ifenthaler.info // xge@ou.edu // rmiller@ou.edu

*Corresponding author

ABSTRACT

Digital game-based learning, especially massively multiplayer online games, has been touted for its potential to promote student motivation and complex problem-solving competency development. However, current evidence is limited to anecdotal studies. The purpose of this empirical investigation is to examine the complex interplay between learners’ motivation, engagement, and complex problem-solving outcomes during game-based learning. A theoretical model is offered that explicates the dynamic interrelationships among learners’ problem representation, motivation (i.e., interest, competence, autonomy, relatedness, self-determination, and self-efficacy), and engagement. Findings of this study suggest that learners’ motivation determine their engagement during gameplay, which in turn determines their development of complex problem-solving competencies. Findings also suggest that learner’s motivation, engagement, and problem-solving performance are greatly impacted by the nature and the design of game tasks. The implications of this study are discussed in detail for designing effective game-based learning environments to facilitate learner engagement and complex problem-solving competencies.

Keywords

Complex problem solving, Motivation, Engagement, Massively multiplayer online games, Game-based learning, Educational game design

Introduction

Proponents have argued that massively multiplayer online games (MMOGs) possess unique affordances to address complex problem-solving skill development that our current educational system is failing to provide (e.g., OECD, 2004) by drawing on a powerful pedagogy: situated learning. It has also been argued that game-based situated learning environments promote student motivation and engagement (e.g., Gee, 2007; Greenfield, 2010). Unfortunately, very few researchers began to move the discussion of complex problem solving beyond descriptive research; the majority of current discourse in the field can be summed up as “games are problems being solved by players [and games are engaging]; therefore, playing games will help [and motivate] people be better problem solvers” (Hung & van Eck, 2010, p. 228). However, this is not sufficient to guide our development of educational games to directly address complex problem solving and student motivation as learning outcomes. In the context of game-based learning, the relationships among problem solving, motivation, and engagement are far more complex than they appear at first.

We argue that complex problem solving and associated cognitive processing and motivational requirements are most impacted by gameplay; and that interactivity captures the most salient features of gameplay as it relates to complex problem solving and motivation. Hence, the purpose of this study was to investigate the interrelationships among complex problem solving, motivation, and engagement in the context of game-based learning and offer an empirically-validated framework that can guide future studies and instructional design efforts.

MMOGs as complex and ill-structured problem-solving environments

It has been argued convincingly that games serve as situated problem-solving environments, in which players are immersed in a culture and way of thinking (Dede, 2009; Gee, 2007). This is especially true for massively multiplayer online games, which are situated in complex and ill-structured problems. For instance, in the McLarin’s Adventures MMOG that served as the testbed for this study, students play the role of researchers who were sent on a mission to explore the habitability of earth-like planets outside of our solar system. The problem is ill-structured because both the given state and the desired goal state are not clearly defined. The desired goal state is vaguely defined as finding...
a planet, on which a settlement area can be built for the humans so that this planet can serve as a colony for the people of the earth. In the first game narrative, the players detect a planet, on which atmospheric conditions (i.e., O\textsubscript{2} and CO\textsubscript{2} levels) allow for humans to breathe comfortably. When they land on the planet, the players can infer from the visibly-apparent characteristics of the surface that the planet resembles a tropical island on the earth.

This problem is also very complex due to the large number of highly inter-connected variables affecting the problem state. This means that changes in one variable affect the status of many other related variables; therefore, it is very difficult to anticipate all possible consequences of any action. Furthermore, not all of the variables may lend themselves to direct observation. Often, knowledge about the symptoms is available, from which one has to infer the underlying state. Dealing with intransparency of the problem variables and the interrelationships among them is often difficult due to time-delayed effects; not every action shows immediate consequences. Hence, complex problem-solving situations often change decremental or worsen, forcing a problem solver to act immediately, under considerable time pressure. Therefore, complex problem-solving situations bear multiple goals, some of which could be contradictory requiring a reasonable trade-off. All of these factors make complex ill-structured problem solving very challenging (Funke, 1991). In the McLarin's Adventures MMOG, players deal with large number of dynamically interconnected variables while they are researching environmental conditions, planning and building the settlement, including planning for sustainable water and food resources, shelters, and so on. Cut scenarios in the game present unforeseen challenges to the players, which change the problem state dynamically, forcing the players to act immediately under pressure.

Eseryel (2006) found that problem understanding and problem solution are not two separate activities in complex, ill-structured problem-solving. Rather, they are intimately connected; they complete each other and develop in parallel. As the solvers try to understand the problem in its entirety by constructing appropriate mental representation to model and comprehend the dynamic interrelationship among problem variables they understand how a change in one variable may affect another, thereby, could mentally simulate in their mind’s eye (Seel, 2001, p. 407) the dynamic problem system in its entirety imagining the events that would take place if a particular action were to be performed. In this way, mental simulation of the solver’s problem space supports causal reasoning during gameplay, allowing one to perform entire actions internally, to judge and interpret the consequences of actions, and to draw appropriate conclusions.

During complex, ill-structured problem solving, it is likely that the solvers either would not possess existing schema of the problem domain or their schema would have to undergo accretion, tuning, or reorganization (Rumelhart & Norman, 1978) to accommodate newly acquired information. The accommodation process is supported by mental models, which are dynamic ad hoc representations of reality to help the individual understand or simplify a phenomenon (Ifenthaler & Eseryel, 2013).

Hence, during complex ill-structured problem solving, solver’s problem representation elicited as a causal representation can serve as a basis for assessing his or her cognitive structure and problem-solving performance (Eseryel, 2006). An individual’s cognitive structure is made up of various schemata and mental models that can be embedded within one another within a hierarchy, which is used to develop procedural knowledge for problem solving purposes within a specific domain (Tennyson & Cocchiarella, 1986). In a recent study, Eseryel, Ifenthaler, and Ge (2013) showed the validity and reliability of the assessment method based on causal representations in the context of complex, ill-structured problem solving during game-based learning. When compared to that of a novice, a domain expert’s cognitive structure is considered to be more tightly integrated and has a greater number of linkages among interrelated concepts. There is thus immense interest on the part of researchers to assess a novice’s cognitive structure and compare it with an expert’s in order to identify the most appropriate ways to bridge the gap to increase problem-solving performance.

Despite the importance of cognitive structure for comprehension, integration of new information, and the ability to solve complex, ill-structured problems (Jonassen, Beissner, & Yacci, 1993), a focus solely on players' cognitive structures is incomplete. Another important prerequisite for successful problem-solving performance during game-based learning is suggested by research on motivational aspects of cognition (e.g., Weiner, 1986), specifically the problem solver's motivation (Mayer, 1998).

This approach suggests that if game-based learning environments can maintain and enhance players motivation, despite the challenges associated with complex, ill-structured problem solving, the players would engage longer in...
gameplay, complete more tasks, which, in turn, will contribute to the sense of competence (Ryan & Deci, 2000; Ryan, Rigby, & Przybylski, 2006). Hence, the longer players are engaged with the game the more their representations of the complex problem scenario of the game resemble to the expert problem representation underlying the game narrative; thereby improving their complex problem-solving performance.

**Motivation and digital game-based learning**

Based on the self-determination theory, the nature and quality of motivation are determined by satisfying three basic needs: autonomy, competence, and relatedness (Ryan & Deci, 2000). Satisfaction of these needs fosters internalized forms of motivation, such as intrinsic motivation (interest), which would lead to higher quality engagement and learning (Ryan & Deci, 2000). Another related key factor that drives motivation and engagement is self-efficacy (Bandura, 1997), which concerns one’s perceived capability for achieving desired outcomes. These key components of motivation are further elaborated in the context of digital game-based learning environments in the following paragraphs.

*Autonomy, competence, and relatedness* are key elements to sustain and maintain one’s motivation (Ryan & Deci, 2000). Autonomy refers to “regulating one’s own behavior and experience and governing the initiation and direction of action” (Ryan & Powelson, 1991, p. 52). For educational MMOGs to be motivating, students should be provided with the paradox of control in an uncertain situation (Csikszentmihalyi, 1990, p. 58). Hence, learners have a sense of control over the environment. Additionally, game-based learning environments should provide learners with opportunities for autonomous choices. In the game-based learning environment, any constrains may limit true choices, which in turn have negative effects on learners’ perceived autonomy.

In addition to satisfying the need for autonomy, self-determined motivation requires that individuals develop a “sense of accomplishment and effectance” (Ryan & Powelson, 1991, p. 52, p.52). Within a game environment, game-players need to believe that they are moving closer to the intended outcome of the game. The challenges they face should match their developed skills so that they can experience attainable challenges, but with some uncertainty of outcomes (Csikszentmihalyi, 1990). Feedback mechanisms within the game are crucial for developing a sense of competence because they will inform learners if they are progressing towards the goal (Csikszentmihalyi, 1990; Fullerton, Swain, & Hoffman, 2004). A number of factors may promote or hinder learners’ perceived competence, including difficulty of tasks and usability of a game (e.g., user interface and navigation features).

*Self-efficacy* is one’s belief on his/her ability to achieve a desired outcome (Bandura, 1997), which has been found to be a good predictor of future learning outcomes (Pajares, 1996). Some factors influencing self-efficacy include performance feedback and social comparison (Bandura, 1993). When people attain their goals, self-efficacy can be increased. People also judge their own ability for a task by observing how other people perform in attaining their goals. In an MMOG environment, players may gain self-efficacy through overcoming various challenges in a game. However, they may lose self-efficacy as they observe other players struggle in the game. When game players have high self-efficacy, they are more likely to put forth effort and be more persistent in pursuing their problem-solving tasks (Zimmerman & Campillo, 2003). In short, when their self-efficacy for challenges of the game is higher, players are more likely to engage in the game.

Finally, self-determination theory stresses the importance of building positive interpersonal relationships in self-determined motivation (Ryan & Deci, 2000). Traditionally, relatedness refers to students’ feeling of belonging in the classrooms, such as acceptance, inclusion, and support. It can also be referred to the quality of the relationship between students and teachers (Reeve, 2006). In a game-based learning environment, relatedness can be extended to the quality of relationships among the players (Ryan, et al., 2006). As players establish a common language and work towards common goals, peer relationships can be strengthened. Hence, a game-based learning environment may foster relatedness as students are engaged in solving complex problems together.

Motivational theories can explain students’ behaviors in a learning environment, such as effort and persistence (e.g. Ryan & Deci, 2000), which are important indicators of engagement. Effort and persistence are directly influenced by the combined motivational factors, such as interest, autonomy, competence, relatedness, and self-efficacy. These motivational factors influence students’ regulation effort and reflection on their understanding of the problem and the quality of solutions (Pintrich, 2000). In a game-based learning environment, effort and persistence can be
operationalized through the amount of time a player spends in a task and the number of tasks that a player accomplishes within a limited time period. In light of our observations of digital game-based learning, problem solving, and motivation, we assume that an in-depth investigation of these complex processes will help instructional designers to understand the relationship between complex problem solving, motivation and engagement and to improve the design of educational games that are most appropriate for learning and problem solving within digital game-based learning environment.

Purpose of the study

Figure 1 depicts the theoretical model stemmed from the literature and represents the influence of motivation and prior problem representation on student engagement, which in turn influences students’ complex problem-solving performance in game-based learning, indicated by their problem representation. The model utilizes self-determination theory and self-efficacy theory (Bandura, 1997; Ryan & Deci, 2000), as well as cognitive structure and problem representation (Ifenthaler & Seel, 2011; Jonassen et al., 1993), as theoretical foundations. Our research model hypothesizes that students’ problem representation can be explained by the interactions among problem representation, factors of motivation, and task.

Specifically, we assumed that the degree of interest (a reflection of intrinsic motivation) during gameplay has a positive effect on students’ engagement (the number of tasks accomplished and time spent playing the game) in the complex problem scenario (Hypothesis 1a). Further, the model identifies a positive influence of the student’s perceived competence during gameplay on his/her engagement (Hypothesis 1b). Additionally, we assume that the perceived autonomy during gameplay has a positive effect on students’ engagement (Hypothesis 1c). Also, the experienced relatedness during gameplay has a positive effect on his/her engagement (Hypothesis 1d). Finally, we assume that degree of self-efficacy for the tasks during the game has a positive effect on students’ engagement (Hypothesis 1e). Overall, we assume that students’ engagement and prior problem representation have a positive effect on his/her problem representation (Hypothesis 2).

Figure 1. Influence of motivation and cognitive structure on problem representation (rectangular shapes depict initial influences variables; oval shapes depict outcome variables; single arrows depict one-way causal relationship in the direct of the arrow, whereas bi-arrows depict two-way causal relationships)

Method

Participants

A rural high school in the Midwest of the United States was used as a testbed for this experimental study. All of the ninth-grade students in the school (ten classrooms) took part in our study. The data reported here were from N = 88 students (50 female and 38 male) from whom we received both parental consent and student assent forms. Their
average age was 14.6 years ($SD = .7$). All of the participants played the game on a frequent basis (at least twice a week).

**Materials**

**McLarin’s Adventures.** *McLarin’s Adventures* is a massively multiplayer online game (MMOG) designed for 8th and the 9th grade students. In this MMOG students are asked to play the role of researchers set in a survivor story where they explore an uninhabited, uncharted island on a distant earth-like planet. The goal of the game is to successfully colonize a planet and become the winning team.

**Complex problem scenario.** The pre- and posttest was identical and involved a near-transfer problem-solving task when compared with the complex problem scenario in the *McLarin’s Adventures* MMOG. Students were asked to play the role of researchers tasked with developing a successful settlement area on a newly found planet, where humans could survive. They were prompted to think what humans needed to survive and were asked to make a list of each of these factors. The solution to the complex problem scenario was represented in the form of an annotated causal representation, which served as the problem representation construct.

**Motivation inventory.** The motivation inventory consisted of the following subscales. Four subscales of the Intrinsic Motivation Inventory (Ryan & Deci, 2000) were used to assess the participants’ interest (INT; 7 items; Cronbach’s alpha = .92), perceived competence (COM; 5 items; Cronbach’s alpha = .77), perceived autonomy (AUT; 5 items; Cronbach’s alpha = .72), and perceived relatedness (REL; 8 items; Cronbach’s alpha = .77) while performing a given activity. Further, the confidence scale (Bandura, 2006) measured the students’ self-efficacy (MCS; 4 items; Cronbach’s alpha = .87).

**Procedure**

Before a year-long implementation of the *McLarin's Adventures* MMOG, the participants received the pretest of the complex problem scenario, which required them to construct their solutions in the form of an annotated causal representation. Then they watched the opening news video announcing the competition by McLarin International to select viable applicants on a space exploration mission. Our pretest data collection (i.e., demographic data and motivation inventory) was introduced at this point as if it was part of the competition in the game to select the applications. A week after the year-long implementation of the *McLarin's Adventures* MMOG, the participants constructed their solutions to the posttest of the complex problem scenario in form of an annotated causal representation and completed the scales of the motivation inventory.

**Analysis**

The motivation variables of interest in this study reflected the change that occurred between the pretest and the posttest of each variable. In other words, we were interested in the predictive relationship between pre-post change in interest, autonomy, competence, perceived relatedness, and learner engagement. The pre-post motivation variables were the difference scores (Posttest - Pretest), with an increase indicating a greater amount of the motivation variable at the posttest and a negative score indicating a decrease of the motivation variable.

Embedded assessment within an immersive learning environment is regarded as a viable method for making inferences on learner’s behaviors (Chung & Baker, 2003). For instance, student engagement during gameplay is assessed by the time spent in the game or the number of tasks completed during gameplay (Reese, Seward, Tabachnick, Hitt, Harrison, & Mcfarland, 2012). Accordingly, the participant’s engagement (ENG) during gameplay was assessed through log-file data operationalized by (a) the number of tasks completed and (b) the time spent on the game.

In order to analyze the annotated causal representations, the HIMATT (Pirnay-Dummer, Ifenthaler, & Spector, 2010) analysis function was applied. The automated analysis function produces measures, which range from surface-oriented structural comparisons to integrated semantic similarity measures. Those measures include four structural
(surface, graphical, structural, and gamma matching, also referred as SFM, GRM, STM, and GAM) and three semantic (concept, propositional, and balanced semantic matching, also referred as CCM, PPM, & BSM) indicators.

Each of the participants’ problem representation (annotated causal representation) was compared automatically against a reference solution. The reference solution included the annotated causal representation, which was used to guide the design of the scenarios in McLarin’s Adventures MMOG, co-developed by a team of expert teachers of various subject matters. HIMATT uses specific automated comparison algorithms to calculate similarities between a given pair of frequencies or sets of properties. The similarity index $s$ for each of the seven measures results in a measure of $0 \leq s \leq 1$, where $s = 0$ is complete exclusion and $s = 1$ is complete similarity.

Reliability scores for the HIMATT measures range from $r = .79$ to $r = .94$ (Pirnay-Dummer et al., 2010). Convergent validity scores lies between $r = .71$ and $r = .91$ for semantic comparison measures and between $r = .48$ and $r = .79$ for structural comparison measures (Pirnay-Dummer et al., 2010).

Based on the HIMATT measures, the dependent variable problem representation was identified as a combination of semantic and structural properties: $\text{PREP} = (\text{CCM}+\text{PPM}) + (\text{SFM*GRM*STM})$. The variable is reported for the pre-test ($\text{PREP}_{\text{pre}}$) and post-test ($\text{PREP}_{\text{post}}$) results. Based on previous research using the HIMATT measures, the aggregation of structural and semantic measures best reflects individual’s problem representation as it includes strong weights of semantic complexity; however, does not neglect the overall structural components (Eseryel, Ifenthaler, & Ge, 2013).

**Results**

Initial data checks showed that the distributions of ratings and scores satisfied the assumptions underlying the analysis procedures. All effects were assessed at the .05 level. As effect size measures, we used Cohen’s $d$.

**Engagement**

Table 1 shows the zero-order correlations among the variables with regard to the first set of hypotheses. Participants’ engagement in the MMOG was negatively related to their change in interest during the game, as was their change in competence during the game. However, participants’ engagement in the MMOG was positively related to their change in self-efficacy while playing the game. Additionally, their change in interest was positively related to their change in autonomy. Finally, participants’ change in autonomy was related to their change in perceived relatedness.

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Engagement (ENG)</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Interest (INT)</td>
<td>-.303**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Competence (COM)</td>
<td>-.257*</td>
<td>.171</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Autonomy (AUT)</td>
<td>.030</td>
<td>.225*</td>
<td>.019</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Relatedness (REL)</td>
<td>.043</td>
<td>-.021</td>
<td>.130</td>
<td>.537***</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>6. Self-efficacy (MCS)</td>
<td>.225*</td>
<td>.109</td>
<td>.038</td>
<td>-.089</td>
<td>-.140</td>
<td>-</td>
</tr>
<tr>
<td>$M$</td>
<td>68.39</td>
<td>-.87</td>
<td>-.89</td>
<td>-.23</td>
<td>-.13</td>
<td>1.78</td>
</tr>
<tr>
<td>$SD$</td>
<td>29.21</td>
<td>1.50</td>
<td>1.88</td>
<td>.99</td>
<td>1.08</td>
<td>19.99</td>
</tr>
</tbody>
</table>

*Note. $* p < .05, ** p < .01, *** p < .001*

A hierarchical regression analysis was used to determine whether the change in motivation constructs during gameplay (interest, competence, autonomy, relatedness, self-efficacy) were significant predictors of engagement (ENG) in the MMOG (dependent variable). Interest (INT) entered into the equation of step one explained a statistically significant amount of variance in engagement, $R^2 = .092, F(1, 67) = 6.76, p = .011$. After step two, with competence (COM), autonomy (AUT), relatedness (REL), and self-efficacy (MCS) also included in the equation, $R^2$
Thus, the addition of these variables resulted in a 13% increment in the variance accounted for. Specifically, participants’ change in interest (INT) negatively predicted engagement (ENG), indicating that despite the loss of interest (INT) during gameplay, the participant’s engagement (ENG) increased (see Table 2). Accordingly, Hypothesis 1a is rejected. Additionally, participants’ change in competence (COM) negatively predicted engagement (ENG), indicating that despite the loss of competence (COM) during the gameplay, the participant’s engagement (ENG) increased. Accordingly, Hypothesis 1b is rejected. However, the participant’s change in self-efficacy (MCS) positively predicted engagement (ENG), indicating that the higher the change in self-efficacy (MCS) during the gameplay, the higher the participant’s engagement (ENG). Accordingly, Hypothesis 1c is accepted. As shown in Table 3, no correlation was found for autonomy (AUT) and relatedness (REL). Thus, Hypotheses 1c and 1d are rejected.

Table 2. Regression analysis for variables predicting engagement

<table>
<thead>
<tr>
<th>Engagement</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interest (INT)</td>
<td>-5.92</td>
<td>2.28</td>
<td>-.303*</td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interest (INT)</td>
<td>-6.23</td>
<td>2.32</td>
<td>-.319**</td>
</tr>
<tr>
<td>Competence (COM)</td>
<td>-3.43</td>
<td>1.78</td>
<td>-.221*</td>
</tr>
<tr>
<td>Autonomy (AUT)</td>
<td>3.07</td>
<td>4.05</td>
<td>.105</td>
</tr>
<tr>
<td>Relatedness (REL)</td>
<td>1.33</td>
<td>3.68</td>
<td>.049</td>
</tr>
<tr>
<td>Self-efficacy (MCS)</td>
<td>.42</td>
<td>.17</td>
<td>.284*</td>
</tr>
</tbody>
</table>

Note. * p < .05, ** p < .01

To sum up, the results indicate that participant’s self-efficacy (MCS) was a significant predictor for their engagement during gameplay (ENG). In contrast, interest (INT) and competence (COM) negatively predicted the participant’s engagement during gameplay (ENG).

Problem representation

In the pretest the participants scored an average problem representation score of $M = .926$ ($SD = .329$) and in the posttest $M = 1.071$ ($SD = .414$). The increase in the quality of problem representation was significant, $t(87) = 3.259$, $p = .002$, $d = 0.347$.

Table 3 shows the zero-order correlations among the variables with regard to the second hypothesis. The participant’s quality of prior problem representation (before gameplay; PREPpre) was positively related to their quality of problem representation (PREPpost) after playing the MMOG. Also, their engagement (ENG) during gameplay was positively related to the quality of problem representation (PREPpost) after playing the MMOG.

Table 3. Descriptives and zero-order correlations for engagement and problem representation variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Problem representation (PREPpost)</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Prior problem representation (PREPpre)</td>
<td>.387***</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>3. Engagement (ENG)</td>
<td>.237*</td>
<td>.051</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>.926</td>
<td>1.071</td>
<td>62.347</td>
</tr>
<tr>
<td>SD</td>
<td>.33</td>
<td>.41</td>
<td>30.18</td>
</tr>
</tbody>
</table>

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

Next, a hierarchical regression analysis was performed to determine whether the engagement and prior problem representation (PREPpre) are significant predictors of the problem representation (PREPpost) after playing the MMOG (dependent variable). Problem representation (PREPpre) entered into the equation of step one explained a statistically significant amount of variance in problem representation (PREPpost), $R^2 = .150$, $F(1, 86) = 15.17$, $p < .001$. After step two, with engagement (ENG) also included in the equation, $R^2 = .197$, $F(2, 85) = 10.44$, $p < .001$. Specifically, participant’s prior problem representation (PREPpre) and engagement (ENG) positively predicted problem representation (PREPpost), indicating that the higher the change in prior problem representation (PREPpre)
as well as engagement (ENG), the higher the participant’s problem representation (PREP_post) after playing the MMOG (see Table 4). Accordingly, Hypothesis 2 is accepted.

**Table 4. Regression analysis for variables predicting post problem representation (N = 88)**

<table>
<thead>
<tr>
<th>Problem representation (post)</th>
<th>B</th>
<th>SE</th>
<th>β</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior problem representation (PREP_pre)</td>
<td>.488</td>
<td>.125</td>
<td>.387***</td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prior problem representation (PREP_pre)</td>
<td>.474</td>
<td>.123</td>
<td>.376***</td>
</tr>
<tr>
<td>Engagement (ENG)</td>
<td>.003</td>
<td>.001</td>
<td>.218*</td>
</tr>
</tbody>
</table>

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

To sum up, the results indicate that both the prior problem representation (PREP_pre) and the engagement during gameplay (ENG) predicted the problem representation (PREP_post).

**Discussion**

The importance of cognitive structure (i.e., problem representation) for complex problem solving is undisputed (Ifenthaler & Seel, 2011, 2013; Jonassen et al., 1993; Shavelson, 1974). However, we argue that cognitive structure explains only part of the problem-solving performance during game-based learning. Another important prerequisite for successful problem solving is the problem solver’s motivation (Mayer, 1998). In the context of game-based learning, motivation literature particularly emphasizes the influence of self-efficacy and self-determination on the quality and outcome of problem solving (Ryan & Deci, 2000; Zimmerman & Campillo, 2003). Therefore, this study was conducted to better understand the influence of motivation and engagement on the learner’s problem representation in the context of game-based learning. Our research model hypothesized that students’ problem representation can be explained by cognitive structure (i.e., prior problem representation) and various aspects of motivation.

Regarding our first hypothesis, we assumed that interest, competence, autonomy, relatedness, and self-efficacy have a positive effect on participant’s engagement. However, interest and competence negatively predicted engagement, indicating that the higher the loss of interest and competence, the higher the engagement (see Table 2). An in-depth analysis of our data and post-hoc interviews showed that students were initially highly interested in playing the MMOG in the classroom setting. However, the game did not fulfill their expectations (e.g., compared to commercial MMOGs), which caused a decrease of interest in solving the task while playing the game. Despite the loss of interest and competence, the participants kept playing the game. This was mainly because the gameplay was the only classroom activity during the class session. On the other hand, we found a significant positive influence of student’s self-efficacy for the tasks on their engagement (see Table 2). Accordingly, students easily overcome the obstacles and tasks in the MMOG, which increased their self-efficacy during gameplay. Accordingly, the results indicated that the increase of self-efficacy led to the increased student engagement putting forth more effort in solving the problem scenarios within the MMOG and being more persistent in pursuing those tasks (Bandura, 1997; Zimmerman & Campillo, 2003). The students’ perceived autonomy and experience relatedness did not influence their engagement.

Regarding our second hypotheses, results indicated that student’s engagement and their prior problem representation had a significant positive influence on their problem representation. Clearly, prior problem representation and engagement were strong predictors of student’s problem representation after playing the MMOG, that is, the higher their engagement and prior problem representation, the higher their final learning outcome. Therefore, it is concluded that one of the critical issues in designing educational games is to sustain student motivation over time during gameplay. Since students are curious beings and like to seek novelty and explore problems, one way to address the issue of sustainability is to provide students with new and challenging game scenarios as they move along in their gameplay tasks to keep them motivated and focused. On the other hand, if students perceive their problem-solving competence increase in MMOG over time, they would be even more motivated and willing to invest more time and effort in the problem-solving tasks (Mayer, 1998; Ryan & Deci, 2000).
Implications

The findings of this study showed that, due to the challenges associated with complex problem solving, student motivation and engagement have crucial impact on students’ development of complex problem-solving competencies in game-based learning. Furthermore, contrary to the current discourse and beliefs of its proponents, the findings of this study suggested that while games may be complex problems to be solved by students, playing educational games do not necessarily lead to improved problem representations. Hence, there is a critical need for empirically-validated instructional design frameworks to leverage the affordances of game-based learning to design effective situated learning environments that can engage learners and support their development of complex problem-solving competencies. The results of this study also implied that such a design framework should clearly articulate on the design principles and approaches to scaffold learners’ motivation and cognitive structures to sustain high-level engagement during gameplay.

Based on the findings of this study and our formal observations during implementation, we argue that there are three modes of interactions that should be carefully designed in educational games to sustain motivation and engagement (Figure 2): (1) interface interactivity, which refers to the direct interaction between players and game systems; (2) narrative interactivity, which refers to the interaction between the players and the storyline; and (3) social interactivity, which refers to the communication and collaboration between human players (see Eseryel, Guo, & Law, 2012 for more details).

![Figure 2. Three-levels of interactivity design for educational games](image)

As shown in Figure 2, all of these interactivity levels should be designed to complement each other in scaffolding learners’ development of complex problem-solving competencies. For instance, realistic immersive interface interactivity is crucial in enabling effective narrative interactivity and in conveying necessary messages to the players. In the McLarin’s Adventures MMOG, a red road was built to guide players’ wayfinding; however, we noted that such an approach may negatively affect players’ sense of autonomy.

Findings of this study also suggested that, in order to sustain student motivation and engagement, it is necessary that the individual tasks are not fragmented pieces of overall complex problem scenario in the game narrative. For instance, in the McLarin’s Adventures MMOG, one of the problem scenario calls for building a settlement area. This is a complex problem. However, in the game, this problem was broken down into discrete tasks, such as calculating the area of the island. We often observed that students, who started playing the game with high enthusiasm, started complaining after a short while, “this is not a game!” Therefore, we argue that traditional instructional design models, which require fragmentation of learning objectives (e.g., remembering a fact, applying a procedure, understanding a
concept), are not appropriate for designing narrative interactivity due to a lack of challenges to intrinsically motivate students.

Findings of this study also suggested that social interactivity during gameplay, such as competition and collaboration with others, plays an important role contributing to learners’ motivation, engagement, and development of complex problem-solving competencies. For instance, the backstory of the McLarin’s Adventures MMOG required players to assume different scientist roles to complete the tasks while playing the game in teams of four. Initially, this led to increased engagement. However, because the tasks did not truly require collaboration among team members with different expertise or roles, the students soon discovered that the tasks in the game had to be completed by each player individually to get game points. When we examined the chat and voice logs of the game, we found that student collaboration noticeably declined as the game progressed; instead a very large portion of the chat content was irrelevant chatter. Hence, we concluded that lack of social interactivity embedded in the game narrative might have negatively affected students’ motivation, engagement, and learning.

Limitations and future work

As with all experimental research, there are limitations to this study, which need to be addressed. First, while our sample size was large enough to achieve statistically significant results, the explained variance for our regression models was rather moderate. This indicates that besides the tested variables other variables may have influenced the outcomes, which were not tested in this study. In contrast to laboratory experiments, design-based research limits the validity due to influences that cannot be controlled outside the study’s intervention.

Second, the issues in game design may have contributed to moderate variance. Thus, our future plans include further investigations when the design of interface, narrative, and social interactivity dimensions are improved per discussion in the previous section.

Third, the implementation of the MMOG needs to be addressed. Our implementation model followed the minimal-external-guidance model where teachers acted solely as facilitators. In contrast, games may also be implemented with a strong emphasis on guidance. Accordingly, future studies are needed to address the effectiveness of different instructional models for implementing game-based learning (Kirschner, Sweller, & Clark, 2006).

Last, assessment of game-based learning is challenging and it is being questioned. Only very few empirical research studies exist that used valid and reliable methods for assessing complex problem solving in game-based learning (Eseryel, Ge, Ifenthaler, & Law, 2011). There is a need for established assessments methods that can be embedded in educational games to determine engagement, progress of learning, and development of complex problem-solving competencies (Ifenthaler, Eseryel, & Ge, 2012a). Accordingly, future generation of educational games should include embedded assessments and provide instant analysis of a wide variety of 21st century skill acquisition and offer personalized feedback to learners (see Eseryel, Ge et al., 2011; Eseryel, Ifenthaler, & Ge, 2011; Ifenthaler, Eseryel, & Ge, 2012b).

Conclusion

This study shows that motivation and engagement in a game-based learning environment have an impact on learners’ problem-solving outcomes. Thus, it is crucial to design game-based learning environment to scaffold students’ motivation and engagement. Not all the games are necessarily designed as complex to engage students in problem-solving tasks; therefore, the assumptions that all games are engaging and that playing games will increase learners’ problem-solving skills are challenged. This study reveals that (1) there are complex interplays between the student engagement, motivation, and problem-solving competencies; (2) the game can enhance or limit learners’ choices and relatedness, which can influence learner engagement; and (3) the design features of the game can affect learners’ self-efficacy and perceived competence. Therefore, in order to foster students’ complex problem-solving competence, educational games should be designed in a way that provides complexity for students to engage in problem-solving tasks, with sufficient autonomy for students to make choices and attainable challenges to help them move closer to their intended goals.
This study was among the first that explored the complex relationships between motivation, cognition, and instruction in the context of digital game-based learning. There are many further questions to be explored in the future; for example, how to design optimal educational games that have the appeal of a commercial game; how to directly capture the data of students’ activities and emotions during gameplay; and how to measure and assess learners’ mental effort, cognitive load, decision-making, and other self-regulation processes. Scientific inquiries along these lines will help us to advance research on complex problem solving in the digital age.

References


Multi-User Virtual Environments Fostering Collaboration in Formal Education

Nicoletta Di Blas* and Paolo Paolini

HOC-LAB, Department of Electronics and Information, Politecnico di Milano // nicoletta.diblas@polimi.it // paolo.paolini@polimi.it

*Corresponding author

ABSTRACT

This paper is about how serious games based on MUVEs in formal education can foster collaboration. More specifically, it is about a large case-study with four different programs which took place from 2002 to 2009 and involved more than 9,000 students, aged between 12 and 18, from various nations (18 European countries, Israel and the USA). These programs proved highly effective into fostering a number of transversal skills, among which collaboration (both remote and in presence), stood out as prominent. The paper will introduce the four programs, the way they were designed to foster collaboration and the data on their impact.

Keywords

MUVEs, 3D virtual environments, Computer-supported collaborative learning, Educational benefits

Introduction

“Most great learning happens in groups” and “collaboration is the stuff of growth”: thus goes Ken Robinson’s thinking as expressed in the successful YouTube video “Changing Educational Paradigms” (Robinson, 2010).

We could not agree more that modern society needs people who are capable of working, learning, negotiating opinions, being creative in groups. But school systems around the world are still modeled on an “individual accountability” basis, which – again, as Sir Ken Robinson points out – depends on the ideology and societal organization of the period in which they were created: the beginning of the 19th century, when enlightenment on one side and industrialization on the other were shaping society in a way never seen before.

It so happens that our children are taught that there is “one and only one answer to a question” (which can be found at the end of the book – and you can’t look it up, that’s cheating!), that there is no room for multiple views, that talking to peers and discussing different opinions is not allowed nor profitable, that in the end what matters is the score the single student gets and there is no point in helping or cooperating with the others. But there is a lot of evidence that these “solo-players” produced by our school systems are not what our new, ever connected, ever interplaying society and professional world needs. The digital natives of today live the majority of their lives in a net of relations, stretching far and wide: the only place where the lines of this net are broken is school. How long can this last?

This paper presents a large case-study with four serious-games programs deployed in the context of formal education, in which collaboration (both remote and in presence) was essential for success. These programs involved thousands of students, from all over Europe, Israel and the USA, interacting, studying and playing over serious subject matters. The results, extensively monitored through a number of means, were quite positive in terms of educational benefits (Di Blas & Ferrari, in press): Participants learnt more about the subjects at stake, gained a number of skills (improved English, technical skills…) and also changed their minds about some issues (e.g., they better understood their peers’ cultures). For the sake of this paper, the most relevant result regards collaboration: the design of the educational experience pushed participants to collaborate, both within the class and their remote peers. The main findings of the research study are: teachers were very keen on organizing the class in such a way as to promote collaboration (e.g., through group-work) and they rated collaboration as one of the main benefits for the students; the students themselves acknowledged they had improved their capacity to work in groups and rated group work and cooperation as two of the most engaging aspects of the experience. These data were confirmed by the online tutors monitoring the meetings in the MUVEs and the interaction in the forums.
Background

Learning and collaboration

Computers in education have a high potential for supporting collaboration. The recent relevant literature about HCI (e.g., Czerwinski et al., 2008; Olsen et al., 2009) and interaction design for children (e.g., Cassell, 2008) shows a pronounced interest in the development of collaborative technologies that can support interaction of groups at a distance or in co-presence.

Theories of cooperation place an accent on the underlying motivation and the dynamics of cooperation. Argyle (1991) suggests that there are three main reasons that can trigger cooperation: external compensation; building relationships and sharing and enriching the activities the participants are involved in. Some authors insist on the intrinsic motivation to cooperate as a key element for successful cooperative learning (McConnell, 2000). Cooperative learning can be defined by insisting on a shared goal in group work (Underwood and Underwood, 1999), or on the relation between collaboration in group work and learning, with a focus on the process (McConnell, 2000). The analytical premises in the field of computer-supported cooperative learning (CSCL) are tied to different theoretical schools. Koschmann (1996) identifies three major theories influential in the CSCL field:

1. constructivism insists on the role of interaction among peers in cognitive development and on the socially constructed nature of knowledge;
2. sociocultural theory, building on the legacy of Vygotsky (1978), places an accent on the role of the tutor, or a more skilled person, and their support in defining ‘the zone of proximal development’;
3. situated cognition theories stress that learning occurs through participation; communities of practice have embedded specific knowledge, and learning occurs by entry and participation into such a community through active sharing.

In the variety of approaches and theories that can serve to explore the relation between interaction and learning in group settings, two common denominators of cooperative learning can be found: the existence of a group, and a shared goal, purpose or outcome to achieve (McConnell, 2000). From here, a rich variety of cooperation patterns may arise and are currently practiced in schools worldwide. For the purpose of this paper, we find McConnell’s framework useful for analyzing the types of collaboration in cooperative learning, based on a set of six aspects: 1 structure (highly structured–no structure); 2 teacher control (high–low); 3 moderation of learning (external–internal); 4 learner motivation (external–internal); 5 learning content (curriculum based–learner based); 6 assessment (unilateral by teacher–unilateral by learner; McConnell’s 2000).

The approach of the programs presented in this paper can be related to constructivism (peer learners discuss and socially construct their knowledge), situated cognition (learning occurs by entry and participation) and sociocultural theory (the teacher acts as a tutor). The shared goal is of paramount relevance, as discussed later in the paper. Eventually, all the aspects of McConnell’s framework are represented, except number 6, in the sense that assessment was “the teachers’ business” and the designers were not involved in it.

MUVEs in education

MUVEs have been used several times in formal education, i.e., situations where students in a class, with a teacher, are given precise learning goals (Dieterle, Clark, in press). We can recall here: Barab’s Quest Atlantis, a persistent virtual world where children engage in curriculum-related quests to save a fantasyland from environmental disaster (Barab et al., 2005; Barab et al., 2009); Dede’s River City (Dede et al., 2005), where teams of middle-school students investigate the social, health and environmental causes of an epidemic in a virtual town; Bers’ Zora (Bers, 1999), a virtual environment used by children with psychological, mental or physical problems to express themselves through the manipulation of virtual objects and characters; AppEdTech (www.lesn.appstate.edu/aet/aet.htm) a graphical MUVE supporting graduate students who work over distance; AquaMOOSE 3D (www.lesn.appstate.edu/aet/aet.htm), a MUVE about parametric equations; MOOSE Crossing (www.cc.gatech.edu/ele/moose-crossing), a text-based MUVE for children aged 9-13; Revolution (educationarcade.org/node/357), a multiplayer role-playing game where students “take part” in the American Revolution as members of a virtual community set in Williamsburg; Whyville (www.whyville.net/smmk/nice), a graphical MUVE designed for children aged between 10 to 12 to communicate with friends, learn math, science, and
history, and build online identities; Critical Life (Rogers, 2011), a MUVE that allows student nurses to practice their clinical skills using the Second Life platform; Virtual Singapura (http://virtualsingapura.com/game/), an intelligent agent-augmented multi-user virtual environment modeled on early 19th-century Singapore, and a variety of artificial intelligence entities called intelligent software agents that act as the learning companions for the learners.

Recent MUVE studies (Aldrich, 2009; Badawy, 2012; Dickey, 2005; Laws et al., 2009; Tobias & Fletcher, 2011) show that a number of important questions have been raised, such as: Do games and virtual worlds work for all learners/subjects? How do we assess learning when it's happening in games and virtual worlds? How does the kind of learning that happens in games and virtual worlds map onto curriculum standards? Authors have a broad understanding of how MUVEs can be designed to support the situated and distributed nature of learning and thinking (Dieterle & Clarke, in press) recognizing the Distributed Cognition, which states that “knowledge and cognition is distributed across objects, individuals, artifacts, and tools in the environment” (Hutchins, 1995), as a contributing theory for 3D virtual worlds in education. As MUVEs are designed to give students problems with several paths to the solution, performance-based assessments, such as proposals or final reports, seem to better assess the pedagogical benefits (Ketelhut, Clarke, & Nelson, 2010). Researchers recognize general benefits for students since 3D virtual worlds can assist with improving self-efficacy (Ketelhut, Nelson, Clarke, & Dede, 2010) and can provide environments that immerse the student in various roles and tasks, encouraging her to become an explorer and experimenter (Rogers, 2011). Many published studies report on the impact evaluation of MUVEs in formal education. For example, authors in (Kennedy-Clark et al., 2009) focus on analyzing the impact of structure in inquiry-learning activities in Virtual Singapura, showing that “adopting a low structure initial activity in inquiry-learning can result in better learning outcomes than using an initial high-structure activity.” Researchers in (Nelson & Ketelhut, 2007) present a review of the emerging use of MUVEs to support interactive scientific inquiry practices revealing that “MUVE-based curricula can successfully support real-world inquiry practices based on authentic interactivity with simulated worlds and tools.” Researchers in (Sancho et al., 2009; Perera et al., 2010) conducted case studies collecting data about motivational aspects of the use of MUVEs in managed learning, i.e., as software tools and digital content specifically intended to support learning activities. Authors recognize that “maintaining student engagement is a major concern in higher education, especially when concepts become more sophisticated and coursework becomes more complex.” Other works report on in-field observation and evaluation studies on collaborative and virtual learning environments, both from a teacher’s point of view, stressing his/her habit changes (Marty & Carron, 2011), and from a student’s perspective (Jong et al., 2010), analyzing the “positive quantitative findings of the study, with a combination of quantitative and qualitative methods of inquiry.”

The case-study

The four programs this paper is based upon are:

• SEE (Shrine Educational Experience); 2002-2004. In cooperation with the Israel Museum of Jerusalem, SEE involved over 1,400 middle and high school students from Italy, Israel and Belgium. It was about the Dead Sea Scrolls and related religious, political, historical issues. www.seequmran.net

• Stori@Lombardia; 2005-06. With the support of the Regional Government of Lombardy (Italy), on the medieval history of the Lombardy region. More than 1,100 students from Northern Italy, aged between 12 and 19, were involved. www.storialombardia.it

• Learning@Europe; 2004-08. With the support of Accenture Foundation, as part of the Accenture Corporate Citizenship investment program, on European history. Since 2004 it has involved more than 6,000 high school students from 18 European countries and the USA. www.learningateurope.net

• Learning@SocialSport; 2007-09. In cooperation with the SBS Master Verde Sport (of the Italian fashion group Benetton), Fondazione Italiana Accenture and the Italian Olympic Committee (CONI) on ethical, social and psychological issues related to sport. Since 2007 it has involved more than 350 young athletes. www.learningatsocialsport.net

All the programs shared the same approach, though of course changes were made over the years following monitoring and evaluation. The basic approach can be summarized as follows:

• Classes underwent collaborative educational experiences supported by Multi-Users Virtual Environments (figure 1)

• These experiences were blended, in the sense that they involved both off-line (more traditional) and online activities. Online meetings (called “sessions”) in the MUVEs were the core of the experience (figure 2).
- The whole educational experience would last between 1 and 2 months.
- There was always an overarching goal: a competition. Anything in the experience mattered to this end, from discussion to homework delivery, from ability games to the quality of interaction.
- Each experience comprised 4 different classes from 4 countries (e.g., USA, UK, Poland and Italy). For each class, two avatars were present in the 3D environment (figure 1). In order to involve more students (and to allow for more in-depth discussion), in year 2007 a third environment was added: a 2D chat, where an additional student for each class was involved, answering difficult cultural questions.
- Each experience was managed by two online guides.
- MUVEs were meant to support interaction and to foster motivation. In MUVEs, games would alternate with cultural discussion and quizzes.
- In the real world, substantial learning would happen (assignments, research, etc.). To start from a fair basis, all the participants were provided with the same set of background content: documents derived from interviews to leading experts (historians, sociologists, etc., according to the subject at stake).

![Figure 1. Learning@Europe’s virtual world, with pictures of the students’ countries](image1.png)

![Figure 2. Overall plan of the Learning@Europe experience](image2.png)
The MUVEs supporting all the programs were developed by HOC-LAB. The current version (WebTalkCube) of the platform is the result of a number of refinements over previous versions. The technical platform is presented in details elsewhere (Di Blas et al., 2012).

MUVEs to support collaboration

Collaboration was one of the main goals of all the programs; it came naturally, since MUVEs are intrinsically collaborative (Bucciero et al., 2011). Making the experience collaborative was a pervasive requirement: almost all of the program’s features were directly or indirectly affected by it. Let us see now some examples of experience’s features meant to support collaboration.

Example 1: The assignments

Students were asked to complete a number of assignments, especially in the Learning@Europe program. First of all, a class presentation had to be prepared before session 1: students had to transform it into an HTML page that would then be shown in the virtual world, together with pictures of the students’ countries. This first assignment required collaboration inside the class. Between session 1 and session 2, a team presentation was required: the four classes taking part in an experience were paired into teams of two classes each, and each team had to collaborate remotely to create a “team presentation”. Eventually, between the second and the fourth (last) meeting, the biggest effort was required; students were asked to:

- make a survey to people in the streets about their perception of their national identity;
- take pictures of monuments, streets, buildings of their town;
- prepare an essay discussing some aspects of their nation’s history (e.g., the influence of religion in the formation of their nation-state);
- prepare in collaboration with their remote peers another essay, comparing their different points of view.

Task 1-3 implied collaboration within the class, task 4 implied collaboration with the remote peers. Furthermore, students within the class were organized into groups for studying the background materials necessary to take part in the online discussions and quizzes.

All assignments thus required collaboration: both within the class and with remote peers.

Example 2: The treasure hunt game

In all the programs, there always was a treasure hunt game, taking place in a labyrinth. The labyrinth was divided into two halves, one for each team. Each team had to look for its own set of objects, following cultural clues. The labyrinth contained a number of boxes: The user had to click on the box, see what the object was and decide whether it was the right one or not. As a consequence to her selection, a piece of a sentence would appear in the middle of the labyrinth: If all the four objects were right, the sentence would make sense. In SEE it immediately turned out that users would play individually and would select any object they found, just trying to make up a meaningful sentence. So both of our goals were not fulfilled: Users did not collaborate, nor did they think about the cultural clue. In the following versions of the programs, rules were changed, exploiting also a technical bug that allowed users to look through the eyes of other users. First of all, it was made compulsory that an object in the maze had to be selected by two users at a time. If one of the players found an object she thought was the right one, she had to call her mate via chat and ask her to look through her eyes so as to select the object together. In addition, points were taken for any wrong selection: In this way, guessing-game was avoided.

Thanks to this changes, the Treasure Hunt was turned into a collaborative game.

Example 3: “Find your way” game

In the “Find your way” game, one user had to move through a maze full of obstacles she could not see; her remote team partner instead could see the obstacles, so she could give her directions, via chat. In addition, if the other team
partners who were discussing in the 2D chat were giving correct answers, the obstacles were made visible (figure 3). This game thus implied a lot of collaboration: each user’s ability and knowledge would contribute to the team’s success.

Figure 3. The “Find your way” game, where a “blind” user moves through a maze with obstacles following a remote partner’s directions via chat

Evaluation

The four programs were extensively monitored over the years through a number of means:

- surveys to teachers, before the experience, after each online session and at the end of the experience;
- surveys to students, before and after the experience;
- session reports by the online guides after each online session;
- forum reports, weekly filled by the online guides;
- direct observations in the classes (4-5 classes per year);
- post-analysis of online sessions (recorded with Camtasia);
- assignments’ evaluation by the tutors.
- focus groups with teachers (20 on average) at the end of each year.

The programs proved effective into fostering a substantial educational benefits of various kinds, from increased knowledge of the subject matter to changes of attitude (e.g., “increase tolerance” towards other cultures, in the case of Learning@Europe). The main results are discussed elsewhere (Di Blas et al., 2009; Di Blas et al., 2012). We shall focus here on data about collaboration. Since over the years the evaluation systems evolved (e.g., scales and questions were adjusted), data will be taken from a specific program, Learning@Europe, in its last year of deployment, when the largest number of participants were involved.

The surveys to teachers (67 respondents) provide evidences of the collaborative activities that had taken place in the class. When asked to describe how activities had been organized, 76.9% of the teachers said that students were organized into groups to complete the assignments and a similar percentage (71.9%) said they had organized their students into groups also during the online sessions. In addition, 53.8% reported that their students had interacted with the remote peers through the team’s forum.

Data about how students worked in the groups are quite interesting. 39.1% of the teachers made the students take turn at the computers, in all other cases they let the students organize themselves according to their skills and
preferences. So, in the majority of cases (85,9%), students “specialized” in the sense that they performed specific roles.

The above data show that collaboration went with specialization: an occasion for all the students to show what they were good at, exploiting their natural talents. This brought about some unexpected benefits, like the rescuing of disaffected or marginalized students (Di Blas & Poggi, 2008; Di Blas & Ferrari, in press).

Collaboration was pervasive: Thus it comes as no surprise that the teachers’ rating of “group work” as educational benefit for their students is quite high: 3,80, on a scale from 1 to 5 (Figure 6).
A teacher reported: “Each one’s skills were resources for the class. They understood that, by playing their role well, the whole team would benefit. I saw none of the usual jealously for those who controlled mouse and keyboard: They stood together, united to win.”

![Graph showing students' improvements in skills](image)

*Figure 6. Teachers’ rating of the students' improvements in terms of skills; group work scores quite high; scale from 1 to 5 where 5 is best, 67 respondents*

Eventually, the relevance of the competition as a group activity also emerged from the surveys: 75.4% of the teachers either agreed or strongly agreed that “competition had motivated theirs students to learn” (23% partially agreed with this statement, 1.6% disagreed and none strongly disagreed).

The surveys to students (535 respondents) shed further light on how much and in what sense collaboration worked. First of all, quite surprisingly group work with class mates turned out to be more engaging than the games (Figure 7).

![Graph showing student engagement](image)

*Figure 7. Students’ rating of the most engaging aspects of Learning@Europe; scale from 1 to 5 where 5 is best, 535 respondents*
Students were also happy to interact with remote peers and rated it as the main reason why they appreciated the online sessions. When asked to self-assess their improvements in skills, students acknowledged they had learnt how to work in groups (Figure 8).

![Figure 8. Students’ self-assessment of their improvement in skills; scale from 1 to 5 where 5 is best, 535 respondents](image)

Exchanging ideas with remote peers was an eye-opening experience for many, like for example a girl who took part in Learning@SocialSport, who wrote in the forums: “In my opinion it is great to meet young athletes coming from different places, who practice sports that are different from mine. There are many differences, due to the different geographical locations, but the great thing is that you can see that a rugby player has an infinite sense of fair play, that a basketball player is willing to pass the ball to his team mate, that a canoeist rows with all her might to make her team win and that a tennis player, even if she plays alone, feels part of a big family. It is the team and the team spirit that unites us”.

The reports by the online guides (79 respondents) pinpoints how chatting with remote peers is the third most successful aspect of the online sessions (being the exploration of the environment first and the promptness in following the guides’ directions the second).

![Figure 9. Online guides’ assessment of the most successful aspects of the online sessions; 79 respondents (on X axis)](image)
The analysis of the forums confirm the above data: socialization is the best-achieved goal (scoring 3.21 – on a scale from 1 to 5 where 5 is best; 39 respondents) among all the goals the forums were supposed to fulfill (figure 10).

![Figure 10. Assessment of the forum’s outcomes by the online tutors; scale from 1 to 5 where 5 is best, 39 respondents](image)

An online tutor reports: “they liked talking about their lifestyle and everyday life. They exchanged opinions about the homework and collaborated in doing it. They were always friendly to each other”.

Eventually, direct observation in the classes revealed that students would group around the computers, cheering, suggesting answers and moves, supporting the players, etc. (figure 11).

![Figure 11. Screenshot from a video taken in a class during a Learning@Europe’s session](image)

A number of collateral benefits related to team-work were detected (during focus groups with teachers and through comments in the surveys): By working on a common task, users can get benefits in the ethical and affective sphere, like increased social commitment (my task is important for my community), sense of responsibility, understanding of deadlines, capability of working in groups and negotiating with peers. Teachers rated 3.93 (on a 1 to 5 scale, where 5 was the best score) the efficacy of “group work” to develop their students’ sense of responsibility. A teacher reports: “My students are learning to take into consideration their peers’ opinions.”
Conclusions and lessons learned

In this final section we discuss some “lessons” on how to foster collaboration in MUVEs’ based educational experiences. These lessons derive from our experience as designers, were refined over the years following our users’ evaluation and were confirmed by the positive outcomes discussed before (for a thorough description see Bucciero et al., 2011).

The reader may note that many of the lessons that follow can actually be applied in other situations, where different kinds of technology are used: it is also the authors’ experience, for example with digital storytelling in formal education (Di Blas & Paolini, 2013).

Lesson 1

*Provide a common, overarching goal perceived as meaningful*

In the case of the above described programs, all the student knew they were involved in a competition in which points would be given and taken for any action, ranging from the quality of the discussion (which meant gaining points) to misbehaviors (which meant loosing points). At the end of the experience, one team would be “crowned” as winner. That a competition can be motivating is known: one of Caillois’ category of game is “agon,” i.e., competition (Caillois, 1961), and “challenge” is one of the 11 elements on the taxonomy of fun by Prensky (Prensky, 2001). As shown above in the evaluation section, the large majority of the teachers agreed that the competition had been a strong motivator for their students. A teacher reports: “during the last online meeting the guide declared us winners: the children roared as if they were at a soccer match. People would come into the classroom asking what all the excitement was about”. There is wide discussion about the relation between motivation, technology, gaming activities and learning (Tran et al., 2012): in our case, we used a (multi-faceted) extrinsic motivation to somehow lure students into studying the serious subject-matters at stake.

Lesson 2

*Split activities into doable tasks*

In a collaborative experience, it is fundamental to split the activity into bitable chunks so that participants feel that they can handle it. Moreover, if there is a competition, participants who are not performing well can feel that there is space for improvement (Di Blas, Garzotto & Poggi, 2009). Surveys had shown that teachers would split the class into groups and that students would specialize in different roles: thus, the “bitable chunks” were the perfect solution to fit this pedagogical implementation. Teachers could assign the different tasks to the different groups/students like in an orchestra, where everyone plays a different instrument thus contributing to the final result.

One example are the assignments in the Learning@Europe program: there were assignments between one session and the following (e.g., preparing the class’ presentation, preparing the team’s presentation…) and the final assignment consisted of different chunks: a survey to people in the streets about their perception of their national identity, a reportage with pictures of relevant monuments, streets, places in the students’ home town, an essay about the students’ own national identity and eventually a collaborative essay about similarities and differences between the two essays by the different classes in the same team.

Lesson 3

*Link tasks together*

Linking the tasks means that the success of one user’s task affects, in some way or another, another user’s task. The more activities are interdependent and coordinated, the more evident the need for collaboration is. This is a low-level design requirement deriving directly from the high-level requirement “enhance collaboration” (Bucciero et al., 2011), and its positive effect is backed up by all the data about students’ collaboration shown in the section about evaluation.
One example is the interplay between the discussion in the 2D chat and the “Find your way” game: When correct answers are given in the 2D chat, the obstacles hindering the player in the “Find your way” are removed. This tight relation between performances inside the team makes it clear that collaboration is essential to reach the common goal.

Lesson 4

Support the expression of diverse talents

A class gathers students with diverse talents: If team-building is the aim, it is highly advisable to design different activities so that every single student can feel her contributing to the team’s success. Surveys show that most of the times teachers let students specialize, that is, take the roles that would best fit them.

Examples from our program are the ability games in the 3D world, typically performed by the “video-gamers” in the class, vs. the in-depth, cultural discussion in the 2D chat, where reading and understanding of the background materials was crucial.

The reader may wonder: Is this “blending” of talents collaboration or rather cooperation? Our answer is that it is a form of collaboration: Participants “play like a soccer team,” where players are assigned different roles, but all deeply influencing one another.

As regards this lesson, a warning must be made: Supporting diverse talents leads to the inclusion of all the students in the experience, not in the sense that everyone is learning the same things but that everyone is involved and tries to do her best. This can be a desirable outcome, for motivation is likely to last beyond the boundaries of the experience and boost the students’ performances; still, it must be kept in mind what was said above: different tasks assigned according to different talents does not mean that the students are all learning reaching the same educational goals.

Lesson 5

Provide (even compulsory) collaboration “sparks”

Collaboration, especially in the frame of formal education where students are under the teachers’ control, is not likely to bootstrap spontaneously: It needs to be triggered, even making it compulsory. This is the case of many activities in our programs where collaboration was made the only way for achieving the final goal. One example is the “Treasure hunt” game. Students were supposed to roam a maze were various objects were scattered, some “right” and some “wrong,” according to a cultural riddle they were to solve. In the first version of the game, it turned out that students did not collaborate at all but each individual player tried to solve the riddles by herself. A new rule was thus introduced: that an object in the maze could only be selected by two avatars together. In this way, students were forced to collaborate and discuss, via chat, whether the object was the right one or not.

Lesson 6

Deal with multi-faceted rather than “square” topics

In all our four programs a lot of collaboration was about the subject-matters: history (medieval history, European history…), national identities, religion and society, ways of living, sport. All these subjects have in common that they provide ground for discussion, comparison, exchange of ideas. Surveys to students confirm how much getting into contact with remote peers was appreciated. Some of the most intense moments of the experiences were the ones in which participants would exchange their opinions and sometimes discover that their ways of living were similar, that their historical roots were common, etc. After discussing about how they spent the week-ends, what music they listened to, what they liked to do in their free time, a French student wrote in the chat: “The Polish are like us!” A US cadet, from the West Point Academy (taking part in the last edition of Learning@Europe) wrote: “Learning@Europe opened me up to a new perspective of history, seeing it from the eyes of people in the countries we study, instead of
just from a book” and “… there’s more to Europe than what we read in textbooks! Other nations have a different perspective on history, which I didn’t realized until working with them.”

Figure 12. The West Point Academy participates to the Learning@Europe program

This fruitful exchange of ideas always took place when the subject at stake was multifaceted: It is unlikely to work with subjects that by their nature have just one side, like for example mathematic, where there is no room for different opinions.

Lesson 7

Give space to the teachers’ contribution

Teachers played a fundamental role: They managed the groups, assigned the roles, checked that deadlines were met, supervised the homework and made sure that online sessions would run smoothly. In a word, they were the experience designers in the class, from a pedagogical point of view. An online tutor wrote: “at one point, the students of class x started paying less attention. I could immediately detect that the teacher had gone out of the class, as the students themselves confirmed”.

Two final remarks must be made. First of all, that what may be called the “novelty factor” very likely did play a role in engaging the students; novelty regarded the overall “package”: Technology, taking part in a competition, collaboration with remote peers, collaboration within the class and even a new relationship between the teacher and the students.

Second, that educational experience based on MUVEs like the ones described in this paper are likely to be special event breaking the school’s routine and requiring a significant effort. They are not – not so far, at least – examples of smooth and full technology integration into the classroom, to support everyday activities: only relatively few schools around Europe “dared” to embark in this demanding activity. This can be seen as a serious pitfall, since it makes the adoption of MUVEs in education difficult, unless a totally different kind of experience (less demanding, more pervasive…) were designed.

Our future research plans include better understanding the relation between technology-based activities and collaboration, trying to shed light on the issue of group work and diverse talents. From a practical point of view, a new MUVEs base program is planned for year 2013, again about history, in which social media for remote collaboration will be introduced.
Acknowledgments

We thank all the people that contributed to our programs: the teachers, the students and the HOC-LAB staff. We also thank our experts (more than 30) who were interviewed to create the background materials, and the pedagogy experts (Thomas Reeves and Michael Orey of the University of Georgia, Athens, USA, and Jennifer Preece of the University of Maryland, College Park, USA), who helped to interpret the evaluation data. We are also grateful to the Israel Museum in Jerusalem and to various sponsors (Accenture International Foundation, Accenture Italian Foundation, and the Regional Government of Lombardy).

References


68


E-learning in School Education in the Coming 10 Years for Developing 21st Century Skills: Critical Research Issues and Policy Implications

Siu Cheung Kong1*, Tak-Wai Chan2, Patrick Griffin3, Ulrich Hoppe4, Ronghua Huang5, Kinshuk6, Chee Kit Looi7, Marcelo Milrad8, Cathleen Norris9, Miguel Nussbaum10, Mike Sharples11, Wing Mui Winnie So1, Elliot Soloway12 and Shengquan Yu5

1The Hong Kong Institute of Education, Hong Kong // 2National Central University, Taiwan // 3University of Melbourne, Australia // 4University Duisburg-Essen, Germany // 5Beijing Normal University, China // 6Athabasca University, Canada // 7Nanyang Technological University, Singapore // 8Linnaeus University, Sweden // 9University of North Texas, USA // 10Universidad Catolica de Chile, Chile // 11The Open University, UK // 12University of Michigan, USA // sckong@ied.edu.hk // chan@cl.ncu.edu.tw // p.griffin@unimelb.edu.au // hoppe@collide.info // huangrh@bnu.edu.cn // kinshuk@athabascau.ca // cheekit.looi@nie.edu.sg // marcelo.milrad@lnu.se // norris@unt.edu // mm@ing.puc.cl // mike.sharples@open.ac.uk // wiso@ied.edu.hk // soloway@umich.edu // yusq@bnu.edu.cn

*Corresponding author

ABSTRACT

One of the curriculum goals of e-learning in school education is to develop learners for 21st century skills through their daily learning activities. This paper aims to discuss the research issues and policy implications critical for achieving such a curriculum goal. A review of literature in the related fields indicates that K-12 schools should take advantage of e-learning to maximize learning opportunities of learners for the development of 21st century skills. We identify six research issues critical for e-learning in school education, namely the realization of developing 21st century skills of learners; the bridging of the gap between curriculum in school and situations in society; the maximization of learning opportunities in the learning process; the collection of evidence of improvement and building awareness of progress; the assessment of 21st century skills; and the provision of teacher development for enculturating learners to develop 21st century skills. We recommend the relevant stakeholders across different countries/regions to consider policies on the goal-setting of curriculum addressing 21st century skills development and bridging gap between school and society; on the availability of digital technology for school education; on the privacy/legal issues of learning data in e-learning process; and on the teacher development for pre-service and in-service teachers.

Keywords

E-learning, School education, 21st century skills, Research issues, Policy implications

Introduction

The advocacy of learner-centered learning and the emergence of digital classrooms lead to the demand for transformation of pedagogical design that supports the development of 21st century skills through domain knowledge learning. We establish an International Research Network (IRN) “Theory and Practice of Pedagogical Design for Learning in Digital Classrooms” under the World Educational Research Association (WERA) for the collaborative endeavors of practice-driven research for building theories that inform and direct the pedagogical design conducive to effective orchestration of learning in digital classrooms.

We anticipate the growing trend toward a more individualized and collaborative learning in school education; where physical classrooms keep its importance in learners’ interaction and socialization, yet learning extending outside classrooms will play a more important role in learners’ knowledge construction. Our goal is to advise researchers and governments across different countries/regions to study related research issues and to develop and implement holistic e-learning policies respectively; and then to spread and deepen the impact of digital classrooms for learner-centered learning in K-12 schools.

This paper aims to share our insights into the critical research issues and policy implications for the effective implementation and spreading promotion of e-learning in school education in the coming 10 years. We first share our thoughts of e-learning in school education from a review of related literature; then discuss the research issues in light of our thoughts of e-learning for facilitating K-12 learners to learn effectively in digital classrooms; and finally draw policy implications that we expect from governments at local and global levels for e-learning in school education.
E-learning in school education in the coming 10 years

The popularity of digital technology drives the use of digital resources and communication tools for learning in school education. To benefit from this type of digital learning, learners need inquiry and critical thinking skills to select and process useful and reliable information from varying sources for learning (Padilla, 2010; Trilling & Fadel, 2009). Learners also need communication and collaboration skills to communicate and collaborate with their peers to complete tasks and share outcomes (Saavedra & Darleen Opfer, 2012; Szewkis et al., 2011). These skills are fundamental components of 21st century skills, which empower learners to successfully acquire knowledge and advance learning in the 21st century (Hoffman, 2010; Rotherham & Willingham, 2009). It is therefore foreseen that the learning goals of school education in the coming 10 years need to address the development of 21st century skills beyond curricular goals of learning domain knowledge.

Both domain knowledge and 21st century skills are the learning outcomes to be concerned in the learning goals for developing 21st century skills. As like domain knowledge, learners’ competency in 21st century skills is also needed to be assessed in the process of classroom learning (Claro et al., 2012; Gut, 2011). The assessment of 21st century skills should therefore be linked with the assessment of domain knowledge, so that teachers can better understand the interrelationship between learners’ gains in these two types of learning outcomes. The related assessment should target at evidence of improvement and awareness of progress, in order to reveal both positive and negative features of learning process (Hoppe, 2007). The digital platforms/tools used in digital classrooms can track and store extensive information on learners’ interaction. Such technological support can facilitate a convenient and systematic record, retrieval and commenting on learners’ information on learning process and learning outcomes. Learners in turn can accumulate learner-centered learning experience and create learner-centered learning portfolio. Teachers in turn can provide timely feedback according to learners’ learning outcomes (Alvarez, Salavati, Nussbaum, & Milrad, 2013).

It is foreseen that in the coming 10 years, the school education sector over the world has to get ready for the creation of digital classrooms which support learners to effectively develop 21st century skills through the day-to-day learning process. Figure 1 describes our thoughts of learning in school education for developing 21st century skills.

![Figure 1. Learning in school education for developing 21st century skills](image)

The introduction of digital resources, digital ways of communication and digital platforms for learning and teaching brings about many opportunities to enhance the learning process in school education in the 21st century. It is foreseen that such learning process is supported by resources in digital and non-digital forms seamlessly inside and outside of digital classrooms (So, 2012), in which learners typically use portable computing devices and social learning networks for the retrieval, selection, and sharing of authentic information from multiple sources (Alvarez et
The learning process for developing 21st century skills is characterized of three emphases (see Figure 1 and Table 1).

Its first emphasis is skills development in both formal and informal learning contexts (Cox, 2013; Huang, Kinshuk, & Spector, 2013). The learners will be engaged in a seamless learning environment to coherently apply various generic skills for in-school teacher-led learning process initiated in digital classrooms and after-school learner-initiated learning process in social learning platforms/tools according to individual needs (Milrad, Wong, Sharples, Hwang, Looi, & Ogata, 2013; Otero, Milrad, Rogers, Santos, Verissimo, & Torres, 2011; Wong & Looi, 2011). The feedback for learners will be in a minimal but sufficient amount for identifying individual needs and directions for future improvement (Caballero, van Riesen, Alvarez, Nussbaum, De Jong, 2014; Sims, 2003; Van Merriënboer, & Sluijsmans, 2009).

Its second emphasis is skills development through both individualized and collaborative learning approaches. On their own or with peers, learners will take responsibilities to apply various generic skills for planning goals, implementing tasks, monitoring progresses and evaluating outcomes in their learning process (Kicken, Brand-Gruwel, Merriënboer, & Slot, 2009; Norris & Soloway, 2009). The feedback for learners will be in a minimal but sufficient amount for identifying individual needs and directions for future improvement (Caballero, van Riesen, Alvarez, Nussbaum, De Jong, 2014; Sims, 2003; Van Merriënboer, & Sluijsmans, 2009).

Its third emphasis is skills development supported by evidence of improvement and awareness of progress. The learning process in the e-learning environments can be designed in a range of activities in authentic learning contexts. Rich evidence of improvement and productive failure could be collected from learners’ performance during the learning process of which can be indications on applying 21st century skills for processing real-life information, reflecting on problem-solving ways, articulating tacit knowledge and negotiating multiple analysis perspectives for knowledge construction (Herrington & Kervin, 2007; Niederhauser & Lindstrom, 2006; Zualkernan, 2006). Learners and teachers would then have many opportunities to look into evidence of improvement and reflection on awareness of progress in the e-learning environments which include all and more than those formative assessments in a continuous manner across the learning process and the summative assessments at particular stages.

Technology plays a crucial role in supporting schools on realizing the desirable learning goals, learning process and learning outcomes as described in Figure 1 and as discussed above. Learners in school education nowadays have many channels to access digital resources and use digital ways of communication to connect with peers for e-learning. This exposes learners to many opportunities for applying 21st century skills in the technology-supported learning process. To prompt learners to be benefited from such learning opportunities, there should be convincing research findings and supportive policy initiatives favorable to the integration of e-learning elements in curriculum delivery, in order to motivate K-12 schools to plan for the curriculum goal desirable for the development of 21st century skills among learners through the learning process in e-learning environments.

The learning process in e-learning environments, as discussed, is expected to consider three key elements for maximizing learning opportunities conducive to 21st century skills development. The first element of blending formal and informal learning approaches helps schools to bridge the existing gap between school curriculum and society situations. The second element of balancing individualized and collaborative learning helps learners to increase awareness of learning achievement on individual basis and also increase motivation to make learning progress with peers. The third element of collecting evidence of improvement and building awareness of progress helps teachers and learners to understand levels of learning outcomes on the formative basis and then make informed decisions on next teaching and learning challenge. Providing learners with personal computing devices with wireless connectivity for accessing digital resources, and providing learners and teachers with digital learning platforms/tools for tracking and storing learning data are technological supports favorable to the realization of these three key elements in the learning process for developing 21st century skills inside and outside of digital classrooms. Research inputs and policy planning in relation to the three key elements and the two types of technological supports are genuinely needed for the process of active, constructive and interactive learning in digital classrooms.

Learners will generate learning outcomes in terms of both domain knowledge and 21st century skills in the abovementioned learning process. Research community needs to investigate the effective ways to assess 21st century skills on top of domain knowledge at different stages of such learning process, in order to support K-12 schools on the appropriate and simultaneous measurement of learners’ knowledge and skills development. The successful implementation of the abovementioned learning process requires teachers to realize the vision of promoting 21st century skills among learners; master the skills in facilitating active, constructive and interactive learning process among learners; master the orchestration of digital and non-digital resources; and recognize the
equal importance of academic achievement and 21st century skills as learning outcomes in school education. The current teacher development related to e-learning has to be adjusted to prepare teachers to transform their beliefs and practice in these four aspects (Kong & Song, 2013). The focus of such adjustment will be placed on the impact of the teacher development programs on teachers’ motivational change and then practice changes for learner-centered learning, but not only the integration of IT use or technological support in the delivery of teacher development programs. One important focus of future teacher development program, as mentioned, is the orchestration of resources in digital and conventional non-digital form for supporting learning and teaching in digital classrooms (Nussbaum, Dillenbourg, Dimitriadis, & Roschelle, 2013). Orchestration refers to how a teacher manages, in real time, multi-layered activities in a multi-constraints context (Dillenbourg, 2013). Considering that teachers fail to adopt many research-based learning technologies, the orchestration view privileges the process of adoption and adaptation in an effort to bring the dilemmas of practice into focus.

Table 1 highlights the technological supports conducive to learning in school education for developing 21st century skills; and summarizes the research issues and policy implications critical for e-learning in school education in the coming 10 years. The next two sections will discuss the critical research issues and policy implications in detail.

<table>
<thead>
<tr>
<th>Table 1. E-learning in school education in the coming 10 years: critical issues</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E-learning in school education</strong></td>
</tr>
</tbody>
</table>
| Learning goals: 21st century skills | • Digital resources  
• Digital ways of communication | • Developing 21st century skills | • Policy for setting up curriculum goals for delivering 21st century skills |
| Learning process:  
(1) Formal and informal learning  
(2) Individualized and collaborative learning  
(3) Evidence of improvement and awareness of progress | • Digital classrooms (personal computing devices and wireless connectivity)  
• Digital resources  
• Development/choice of digital learning platforms/tools | • Bridging gap between school and society for linking curriculum content to real life  
• Learning models for developing 21st century skills  
• Performing learning analytics for collecting evidence of improvement and building awareness of progress | • Policy for bridging gap between school and society for linking curriculum content to real life  
• Policy for availability of personally-owned computing devices and free wireless connectivity for developing 21st century skills  
• Policy for addressing privacy/legal issues in e-learning process in which learning data are being tracked |
| Learning outcomes: 21st century skills assessment | • Platforms/tools development for assisting assessment of 21st century skills | • Assessing 21st century skills | |
| Teacher development:  
(1) Learner-centered learning  
(2) Active/Constructive/Interactive learning | • With or without platforms/tools development for teacher development | • Providing teacher development for facilitating learners to develop 21st century skills  
• Scaling up those sustainable teacher development models | • Policy for teacher development for pre-service and in-service teachers |

**Research issues**

Six research issues are considered critical for realizing the changes in the goal, process and outcome of learning in school education in the coming 10 years. The first research issue relates to the realization of learning goal of developing learners for 21st century skills. The research community can contribute in this regard by conducting evidence-based research which concretely identifies factors that motivate school leaders to change school-based planning priority with regard to the development of 21st
century skills. These research efforts aim to promote school leaders to truly recognize the benefits for K-12 learners who master 21st century skills; and so to take action to give priority to the development of 21st century skills when drawing up school-based plans for e-learning integration into curriculum delivery, as well as solicit campus-wide consensus on this planning direction for a smooth curriculum implementation.

The second research issue relates to the bridging of the gap between schools’ curriculum delivery and society’s real-life situations. In the knowledge- and competency-oriented society, learners are demanded on applying knowledge across different subject domains for solving day-to-day problems in the home and community environments. The research community can contribute in this regard by promoting K-12 schools to meaningfully link the content of school curriculum delivered to learners and the application of domain knowledge acquired by learners. One of the promising directions is promoting the creation of digital classrooms in K-12 schools to expose learners to the authentic and contextualized learning environments for promoting cognitive engagement, from the perspective of “cognitive realism”. “Cognitive realism” in the field of education refers to the process of immersing and engaging students in the use of simulation materials for complex learning tasks for “realistic problem-solving” (Herrington, Reeves, & Oliver, 2006, 2007). The research community could help motivate K-12 schools to value the achievement of “cognitive realism” in subject learning, with an emphasis on learning scenarios with multiple outcomes and learning tasks on higher-order thinking, and thereby motivate teachers to select appropriate digital resources for supporting learners, without teacher mediation, to explore, think, decide and perform the steps that an expert would take to solve complex problems of the subject topics.

The third research issue relates to the realization of learning process of developing learners for 21st century skills. The successful promotion of learners’ 21st century skills demands the paradigm shift to learner-centered learning in digital classrooms. It concerns the adequacy and relevance issues of curriculum delivery, in order to allow learners to adequately access resources and scenarios that are related to real-life problem-solving when they develop 21st century skills. Teachers therefore need to change their teaching motivations and pedagogical practice for leading learners in two ways: engaging in active, constructive and interactive learning process for their application and reflection of 21st century skills in school education, and raising their social awareness among the learning processes in school, in community and in home environment. The research community can contribute in this aspect by investigating e-learning models that inform K-12 schools of the promising approach to technological use and pedagogical designs at classroom level for blending formal and informal learning approaches, creating authentic and contextualized learning environments, and balancing individualized feedback provision and self-directed learning process.

The fourth research issue relates to collecting evidence of improvement and building awareness of progress for supporting pedagogical decision-making of teachers and maximizing learning opportunities of learners in developing 21st century skills in the learning process. The advocacy of learners’ active learning and the trend of digital classrooms using various learning management systems and social network platforms place new demands on teachers for transforming pedagogical practices in school education. Data mining techniques are designed commonly for collecting large-scale of data, extracting actionable patterns, and obtain insightful knowledge (Long & Siemens, 2011). The integration of data mining techniques into the learning analytics platform allows for a systematic and real-time approach in identifying effective pedagogical changes (West, 2012). However, the information from learning analytics have not yet well used to inform learning and teaching due to various reasons (Ferguson, 2012). One of the reasons is that a good part of the data accessible to learners and teachers may not be able to inform pedagogical decision making because large-scale assessment data are usually not linked to learners’ learning processes and outcomes. Another reason is that the technologies used to track learners’ learning usually focus on activity tracking instead of knowledge tracking, which do not guide learners and teachers to identify learning problems and make right decision-making to address these issues. Therefore, efforts are needed for researchers to explore ways to provide learning analytics for collecting evidence of improvement and building awareness of progress among learners for appropriate pedagogical decision-making in e-learning environments in school education.

The fifth research issue relates to the assessment of learning outcomes of developing learners for 21st century skills. As both domain knowledge and 21st century skills are the learning outcomes to be concerned in digital classrooms in school education, there should be a coherent and formative mechanism for the assessment of 21st century skills in addition to domain knowledge. The research community can contribute in this regard by studying the ways to support schools to use IT for assessing 21st century skills with a link to domain knowledge.
The sixth research issue relates to the teacher development for facilitating learners to develop 21st century skills. The research community can make corresponding contributions in two ways. First, academics can offer teacher development programs which introduce teachers to the latest development of theoretical frameworks and practical strategies on promoting 21st century skills in e-learning environments, and provide teachers with support ranged from full-scripting scaffolds for less competent teachers to less-scripting scaffolds for more competent teachers. Second, academics can investigate the scalable and sustainable teacher development models that foster teachers’ capability to promote learners’ 21st century skills. The models are expected to facilitate first the sustainable collaboration between K-12 schools and research community for longitudinal studies that enable teachers to iteratively design, apply, reflect and refine their pedagogical practice for realizing learner-centered learning in digital classrooms; and then the scalable sharing among teachers in K-12 schools within the same school and then across the school community that teachers lead professional development on their own for enhancing competency in orchestrating teaching and learning in digital classrooms.

**Policy implications**

Building on the above discussions, we recommend four types of governmental policy supports, at local and global levels, for the sustainable and scalable promotion of digital classrooms in school education in the coming years.

The first type of policy supports relates to the set-up of curriculum goals which bridge the gap between school and society for linking curriculum content to real life, for delivering 21st century skills. K-12 learners nowadays are expected to develop both disciplinary knowledge and 21st century skills through day-to-day learning in digital classrooms. Schools are driven to set curriculum goals which emphasize the inculcation of both disciplinary knowledge and 21st century skills. They are also driven to connect school curriculum with society environment; in which schools link learners’ formal learning under the curriculum implemented in school and their informal learning echoing community’s present situations, and at the same time make campus learning environment transparent to parents and the community. Governments should therefore make policy that officially announces a curriculum emphasis on 21st century skills at national/regional level; and then provide schools with policy incentives to motivate them to prioritize 21st century skills high in e-learning plan and curriculum delivery at school level, re-interpret the school curriculum for a link-up with 21st century skills, identify the possible niches for the integration of 21st century skills into curriculum delivery across different subjects, make curriculum planning with attention paid to its accessibility to sufficient resources and its relevance to real-life situations for skill development, and make long-term school-based plans that indicate their overall directions for curriculum implementation in digital classrooms.

The second type of policy supports relates to the availability of personally-owned computing devices and free wireless connectivity for developing 21st century skills. As discussed, digital classrooms in seamless learning environments enable learners’ development of 21st century skills. Learners therefore need infrastructure supports on the engagement in 1:1 learning environments and the acquisition of stable Internet connectivity both on campus and at home. As the current school learning environment is not sufficiently digitalized, governments should provide schools with flexible policy initiatives for collaborative work on realizing the progression from physical classrooms for traditional knowledge transmission to virtual classrooms for authentic and contextualized learning. Examples of the flexible policy initiatives include funding schools to build IT infrastructure and wireless network for the creation of digital classrooms on campus; mobilizing parents to equip learners with personally-owned device for school learning in the 1:1 learning setting in digital classrooms; and subsidizing learners to own personal computing devices and network connectivity at home for seamless learning extended from digital classrooms.

The third type of policy supports relates to the concerns on privacy/legal issues in e-learning process in which learning data are being tracked. The related policy supports can address the trend of evidence-based approach to learning analytics, so as to facilitate the collection of evidence of improvement and the building of awareness of progress among learners. Learners in digital classrooms learn with digital forms of resources and digital ways of communication. This learning process can be facilitated by one-stop digital learning platforms which support learners’ convenient connection with learning resources and learning peers, as well as systematic recording and tracking of learning process. Governments should therefore support K-12 schools on the deployment of digital learning platforms which have learning tracking capability and consider data privacy protection in two ways: guiding K-12 school practitioners to select existing platforms of a similar kind for school use, and encouraging the IT-related
industry to tailor-make digital learning platforms for K-12 schools in their own countries/regions. As learners’ learning processes are recorded and retrievable on those digital learning platforms over a long period of time on a cross-platform and cross-country basis, this raises a national and even a global concern on protecting the privacy and copyrights of learners’ learning information. Governments should make a policy, at national and global levels, on monitoring the storage, retrieval and destroy of information on learners’ learning processes in this regard.

The fourth type of policy supports relates to teacher development for pre-service and in-service teachers. In digital classrooms which are conducive to learner-centered learning, teachers are expected to act as facilitators who observe learners’ learning processes and provide timely feedback. This role would be relatively new to most teachers who get used to the role as learning authority in traditional teacher-centered paradigm for controlling learners’ learning in classrooms. This drives the need to empower teachers with the capability to act as learning facilitators in digital classrooms for creating e-learning environments and design e-learning activities that promote learners’ authentic and contextualized learning. Governments should therefore support teacher training institutions and K-12 schools to provide pre-service teachers and in-service teachers, respectively, with sustainable and scalable teacher development for the continuous transformation and implementation of teaching practice desirable for digital classrooms.

**Conclusion**

E-learning in school education in the coming 10 years has the goal of developing both domain knowledge and 21st century skills among learners in digital classrooms. This paper introduces a vision of learning in school education for developing 21st century skills; and discusses six critical research issues and four policy implications for promoting e-learning in K-12 schools.

The successful achievement of e-learning vision as discussed in this paper needs the joint effort of three related stakeholder groups. First, policy makers should put effort as discussed for charting the way forward for wider and more effective adoption of e-learning in K-12 schools in their schools districts/states/regions/countries. Second, the research community should work for the critical research issues for supporting the development and dissemination of theoretical-based e-learning pedagogies or resources. Third, the practitioners in the field, including school leaders, teachers, parents, learners and business partners, should put effort in realizing the e-learning vision as discussed in this paper — school leaders need to steer goals and directions of e-learning programs; teachers need to develop and implement sound e-learning pedagogical practice; parents need to acquire digital devices and e-learning resources for learners’ seamless learning after class; learners need to learn with diverse subject-related digital resources that encourage active engagement in constructive learning and peer interaction for developing domain knowledge and 21st century skills; and business partners (including IT sector and content providers) need to provide schools with supports on technical settings of hardware and software, as well as curriculum-based e-learning resources. These joint inputs altogether help for the progressive realization of learner-centered learning in digital classrooms among K-12 learners in their school education.

**References**


Changing Pedagogy for Modern Learners – Lessons from an Educator’s Journey of Self-Reflection

Terry Taylor
School of Computing and Information Systems, Faculty of Science and Technology, Athabasca University, Canada // terryta@athabascau.ca

ABSTRACT

This paper describes the progressive development of some of my current attitudes, beliefs and methods in education, specifically how I believe the modern learner differs from the learner imagined in traditional teaching models. It further explores some ideas and real world attempts to better serve the needs of the modern learner by emphasizing process rather than content.

Keywords

Case study, Evolution in education, Interactive technologies, Reflection

Introduction

I was born in 1952, and as a consequence of that circumstance, my education was a traditional one. In grade school we were taught to write in our notebooks on alternate lines to facilitate corrections. We re-copied endlessly in the ultimately hopeless pursuit of some kind of perfection...in grammar, in punctuation, in the very form of the letters. We drew endless spiral tunnels to perfect the curvature of our penmanship.

Math was done with a stubby pencil on a much erased piece of paper, or, worse, on a dusty blackboard under the watchful eye and ever present yardstick of Mr Buleski, who paid out in brisk taps to the anatomy for every error, chiding and deriding the feckless un-prepared, or simply confused child.

We learned by rote, by repetition, by memory, by recitation in lockstep. Research was dusty library books, through which I could often be found flipping madly, looking for anything that might be remotely relevant to the topic at hand, sweating out paragraphs, themes, essays and term papers. In composition, the ‘outline’ was a straitjacket to any creativity that may have tried to express itself, placing us on a narrative path as straight and unyielding as a railroad track. Linearity was all.

Throughout my school and university education, nothing much changed. The teachers grew less inclined to strike us as we grew older and as educational thought slowly evolved away from the use of fear as a motivator. I sat in huge lecture halls instead of small classrooms, but the pedantic drone of the lecture, the rigidity of the curriculum and the mind numbing absence of creativity, imagination or vision never changed. The night before a final would find me applying my considerable ability to memorize facts to some purpose which even then, and even more so in retrospect, seemed breathtakingly pointless. Memorize the Latin and common names of the fishes of Alberta and their distinguishing characteristics, attend the exam on two cups of coffee and no sleep, regurgitate the mental meal of the previous night like the winner of a pie eating contest, then set about forgetting the whole experience.

Learning was not dynamic, it was not interactive, it was not intuitive and it was certainly not in any way self-directed.

When I finished my undergraduate degree and commenced my studies in Education, I began to learn about the methods and values of the profession from the other side. Looking back, it seems to me that what we were mostly taught was how to manipulate learners, how to lead the horse to water and then shove the hose down his throat and start pumping in the H2O. Never mind if the horse was thirsty, never mind if perhaps he may have preferred a nice Chablis, or a cup of tea.

It wouldn’t be until 1979 that I would first hear the epic words “We don’t need no education, we don’t need no thought control. No dark sarcasm in the classroom, teacher, leave those kids alone” (Waters, 1979), but as I left High School and entered University in 1970, students and other young people on campuses and on the streets all over North America, and in Europe, had begun to question the most fundamental political, economic, cultural and moral
assumptions of their parent’s generation. Also around that time, the first hints of what was to come in the future of computer technology began to appear. Bills had already been turning up in the form of punch cards for a few years at that time, (do not fold, spindle or mutilate!) but computers were still mostly only found in the back rooms of large corporations and government agencies.

Then, one memorable day, I saw “Pong” for the first time.

In one sense it was just incredibly bad television. The little square bounced back and forth like a ball on a tennis court, the white ‘paddles’ moved up and down, blocking and returning the “ball”, or not (SCORE!). Just watching this would of course hold little interest…but…..you could control the paddles. You could alter their size, and the speed of the ball. In other words, it was interactive, and that, though I may have only dimly grasped it at the time, changed everything.

Not long afterwards I had my first teaching job in a rural elementary school, and my first hobby computer, a Texas Instruments TI-99, with 16 K of RAM. It connected to the television, and stored data in analog on a portable cassette tape recorder. By assiduously copying two or three pages of BASIC from the pages of a hobbyist magazine, I could make a tiny stick man walk back and forth on the screen…take that, Pong!!!

In my elementary school classroom I strove not to be Mr Buleski. I wished to find and develop in each of my students the unique talents and interests I knew they all possessed. Since I was dealing with a split grade class (this was a 4 room K-9 school), I had to juggle two lessons at once (give this group some busy work, teach the other group for a while, switch).

It occurred to me that I should develop “learning stations”, areas in the classroom where activities and resources for a certain subject could be arrayed, and among which students could move somewhat freely, within certain constraints aimed at ensuring that each did an acceptable amount of each of the subjects deemed appropriate by our curriculum.

Instead of 25 students sitting in neat rows, reciting in unison, listening in stuporous boredom, or performing repetitive individual drills, I had small knots of students clustered at 8 or 10 learning stations about the edges of the room. I encouraged students to place their desks where they wanted…in a group with their friends, or off by themselves as they preferred. I had a large library table near the blackboard for those occasions when I wanted a group of any size to attend to something I wanted to tell or show them.

My principal, perhaps twice my age and a product of the English public school system, was dismayed and not particularly supportive, but, much to my surprise, a somewhat forward thinking curriculum specialist from the County not only defended but actually encouraged me in this approach, providing learning resources that worked well for individualized learning, and referring me to sources on “modern” educational thought which validated what I was trying to do.

The kids LOVED it. Kids in other classes were envious. Parents, after the predictable rural skepticism about anything new-fangled, realized their kids were ENJOYING school and also became supportive.

I would like to say that I continued this approach throughout my 6 years of teaching in the public system, but the sad truth is, I eventually burned out on the huge extra workload the system created for me and went back to more conventional approaches, though I think I can say with some pride that I never did become Mr Buleski.

I left teaching in 1980 to pursue other interests…first music, and then, when faced with the realization that I was not prepared to spend my life in vans, converted school busses, bars and hotel rooms……computers.

It was while working as a customer support person for one of my cities first consumer computer stores that I became involved in working with PCs for achieving practical goals (i.e., something beyond making a little stick man walk back and forth on the screen), and suddenly found myself teaching again, only this time I was teaching adults how to use these exciting and powerful new tools.
These learners were self-motivated, they were, in the parlance of the time “turned on”. Their individual backgrounds, needs and goals in learning were unique and varied and it seemed like new tools and applications were being developed daily. Around that time I read “The Third Wave,” and I realized that, as Pong had promised 10 years earlier, everything had indeed changed.

**The chainsaw analogy**

By 1983 found myself back in an actual classroom, with adults this time, working in a small college, and teaching, among other things, computer applications and what we then called ‘computer literacy.’

One day during my first term I happened to come upon a young woman working on a term paper, using a word-processing program as I had encouraged my students to do when writing for other courses. Beside her keyboard lay a sheaf of pages, which I realized were her corrected final draft….handwritten.

My mind flashed back to the Language Arts notebooks of my youth, with the writing on every second line and the corrections and additions on the alternate lines, and to the painful experience of re-copying, sometimes repeatedly, to get that perfect final draft, and I had at that moment what can only be adequately described as an epiphany.

I had grievously failed this young woman, and my other students. I had shown them the features, and demonstrated which buttons to press, but I had failed to explain the basic POINT.

Sometime later I would use the analogy of an old-school lumberjack given a chainsaw to replace his axe. If all you told him was that he could cut down 20 times as many trees with this tool, he might not at first succeed. Only when he discovered, or was told, that you had to pull the starter cord first would he begin to get the results the tool was capable of delivering.

Out of that came a growing awareness that I had to not only teach my students the mechanics of using the new software tools, I had to also get them to think and work in different ways.

I began to introduce Word Processing to students by telling them that, first and foremost, they had to change the way they approached the task of writing.

No more writing on the alternate lines, so to speak.

As I refined my understanding of what made writing with a computer different I began to see that the real power of the word processor lay in the fact that it allowed us to separate the mechanical and cosmetic aspects of writing from the creative, to treat them, as they rightly should be, as two basically unrelated processes.

Let us visit, one last time, that 8 or 9 year old writing his theme in a classroom in 1960. He does not want to be doing this. He would rather be on the playground. He is not gifted with naturally tidy handwriting, so the production of something legible is a challenge. He is thinking about punctuation. He is thinking about grammar, and spelling. He is slavishly following a corrected rough draft on the adjacent page of the notebook. The draft was itself written by slavishly following a printed outline. This is the third, or possibly the fourth time he has written this piece.

He does NOT wish to write it again.

Is this young writer’s creative fire stoked up? Is he writing from his heart, from his soul, or even from his brain? If he has a sudden inspiration, will he risk all to alter the content of his composition?

Of course the answer to all the above is “No.”

Back in 1983, I began to ask my students to consciously attempt to separate the creative process from the rest of the job of writing. I encouraged them to try and free themselves from the notion that they had to start at the beginning and end at the end. I told them not to worry about grammar, punctuation or spelling, or even about the logical
sequence of ideas that would make the final version of their composition flow smoothly. I forbade them to develop handwritten outlines or drafts and transcribe those to the computer.

“Write as if you were talking to a friend,” I told them. As fast as the ideas come, let them out onto the screen. Deal with the rest…the mechanics, the cosmetics…later.

Years later I would hear the term “living document” applied to what one is working with in a Word Processing program, and wished I had thought of that!

What makes modern learners different?

Over time, and as people’s life experience with computer technology became greater, each succeeding generation became more intuitive about the way computers can be used. My children literally grew up with computers and computer games. They think differently than I did as a young person. They multi-task, they are non-linear, they value concepts over data (data they take for granted…it’s everywhere). Learning for them does not mean acquiring facts, it means acquiring insight. They don’t need to KNOW, they need to UNDERSTAND.

Since an increasing number of activities and occupations are heavily dependent upon technology, which is itself changing rapidly, “facts” are, at best, a steadily devaluing currency…in many disciplines, today’s ‘facts’ are obsolete and irrelevant tomorrow.

I remember suggesting to a Historian once that he was lucky to work in a discipline where the base of knowledge was not constantly shifting under his feet, as it does for those of us in fields like Information Technology (IT) or Computer Science (CS). He was quick to disagree, pointing out quite convincingly that his area of interest was also dramatically affected by changing technology, and that even the presumed ‘facts’ of historical occurrences are not, as we say, carved in stone, even those which may literally and actually be carved in stone.

Access to information has changed dramatically as well. Ask me any question, anytime, anywhere, and all I need to do is pull out my smartphone, which is really a tiny, mobile computer directly connected to the largest repository of human knowledge ever to exist, and I can, in seconds, give you a detailed and likely quite reliable answer. Not all information on the Internet is accurate of course, but being able to check a variety of sources in a very short time means we have a pretty good shot at finding valid answer to almost any question.

For example, the answer to the question “What is the meaning of life, the universe and everything?” is, as we all now know, “42.”

Levity aside, it is undeniable that access to facts no longer means cramming them into our brains. In the world we live in, the ability to find information, to evaluate it, to synthesize, process, interpret, document and communicate it, have become the relevant qualities of a successful individual.

Today’s learners are trying to prepare themselves for a future that is not at all clearly seen. The pace of change is such that doctrine has little value and adaptability is the key quality for success.

Today’s learner demands relevance, and expects innovation, flexibility and the opportunity to develop those skills that will give them the ability to adapt, to learn independently (for learning is now, more than ever, a lifelong process), and to collaborate effectively.

To properly serve today’s learner, educators must facilitate, empower and support, and institutions must create open and flexible learning environments.

Technology has both driven and enabled these changes, and, just as we learned to write differently (and arguably much better) by using word processing software effectively, we are also learning to teach and learn differently as we gain insight into the true power of the tools technology provides.
An example of the new pedagogy applied

In the late 1980s I began teaching introductory computer courses at a small Native college program operated by the Yellowhead Tribal Council, a program which has since grown into Edmonton’s only First Nations College, the Yellowhead Tribal College, with which I continue to maintain a relationship to this day. The courses were from Athabasca University, and over time I began doing more and more seminar instruction for AU at YTC and other cooperating institutions, eventually becoming a full time Course Coordinator and Tutor in the Schools of Business and Computer Science at AU, positions I continue to hold today.

Athabasca University has been delivering courses online for many years now, but in fact one of the first AU courses to be offered online was one of my courses, COMP 200, a broad survey course which we use as an introduction to CS topics for our full-time students. I wrote the first online version of COMP 200 under the guidance of my colleague, friend and mentor Dr Peter Holt back in the early 90’s and it has been a cornerstone of our CS degree programs ever since.

In that course we experimented with the development of our own online texts, beginning around 1990 and continuing for several years in what proved to be an interesting but ultimately unworkable approach. The problem was that the maintenance of currency in such a text was an impossible task for one person to keep up with, amongst the other duties and responsibilities of my position. This was of course well before ETexts became a widely available option from publishers.

The idea of using primarily online resources instead of a conventional printed text was a compelling one however, especially as the number and quality of websites containing relevant information grew rapidly.

Publishers eventually responded to this expectation in the market and today the availability of EText versions of University texts is almost a given. For some courses and topics however, it may be difficult or impossible to find any single text that adequately covers the scope of content needed.

When I was approached to develop a new course on Interactive Technologies and Human Computer Interaction (IT and HCI), it was obvious from the start that this course would present exactly that challenge. For one thing the breadth of the topic is enormous…it touches on all aspects of HCI, which is to say most aspects of computing.

Another factor was the pace of change within the field, which was fast even by the standards by which we reckon rates of change in computer technology generally. The whole field of HCI is itself relatively new, and not even uniformly defined as yet, if indeed it ever can be.

In addition, I wanted to explore the possibilities of allowing students a degree of control that I at least had never seen given in a course before. I wanted students to be able to make choices, not only about WHAT they learned, but also about HOW they learned, and what they produced by way of evaluated deliverables.

Our goals for the new course, to be called Interactive Technologies, were to provide a broad review of all facets and elements of interactive technologies and Human Computer Interaction, under four main topic areas: the context of HCI, input and output, application interfaces and design/programming. The course was to be exploratory in nature, and designed in such a way as to ensure currency in a fast changing field. A more detailed overview of the topical content may be found in the appendix.

Every activity of course had specific learning objectives identified and stated, however, as the notion that learning should have a clear goal is one aspect of traditional pedagogy I am NOT about to abandon.

My belief was that, for this subject area at least, the learning needs of students would best be met if they were able to get a big picture view of the subject by looking at a series of specific topics in defined areas, and that exactly which specific topics they looked at did not really matter.

There would need to be parameters of course. Hearkening back to my ‘learning station’ experiment in my grade three-four split class of 1975 I decided I could use a similar evaluation system, requiring students to earn a certain number of assignment ‘points’ in each of the main topic areas I had identified for the course.
Within each general topic area I then developed as many relevant learning activities as I could think of, each involving research, analysis, interpretation or synthesis, and reporting. Some actually involved a hands-on activity such as the use of a Rapid Prototyping tool or a basic programming interface.

The choices available to students did not end with being able to pick which activities they undertook. The activities themselves were designed to be open-ended and flexible as well. For example, one activity might ask students to profile a prominent HCI developer. This gave students the opportunity to, first of all, scan the general information available online to find out who fit the bill. Already, just in the process of selecting subject for their discussion they were learning about the people who shaped the way we interact with computers. Once the subject of their exploration was chosen, they would access a variety of resources discussing that person’s work, from which they would glean not only the material for their report, but also, along the way, be exposed to information about the context of the work, who the developer collaborated with, and what the state of the specific technologies they worked on was before, and after their work.

Students were free to use whatever type of learning resources suited their own learning style….they could read, listen to online lectures, watch videos or presentations, discuss the topics in online forums or access the information in whatever form they chose. They were also free to proceed through the sections of the course in a non-linear fashion if they chose. No single topic area was ‘first’ or ‘last’ in the course, so they could start anywhere.

Flexibility and choice was also extended to the format of the material they submitted. Most of the activities could be done as an illustrated article, a PowerPoint presentation, a website, a video or an audio speech. A few, such as an activity using MIT Scratch (Scratch, n.d.) did have a prescribed format for the submission, but even there, choice existed in terms of what sort of program they built.

Evaluation was based on a certain number of possible points being available for each activity, based on the difficulty of the task. Activities were assigned ratings for difficulty levels of Easy, Moderate and Challenging. There was also a set of success levels set for the activities, allowing the person evaluating them to rate them as Basic, Complete or Exemplary.

An “Easy” activity would earn 2 to 4 point depending on the ‘grade’ the submitted material received; 2 for a Basic rating, 3 for a Complete, and 4 for an Exemplary.

A “Moderate” level activity received more points, i.e. 3, 4 or 5, and a “Challenging” activity received 4, 5 or 6.

Students were told to plan their activities so that they would be targeting a minimum of 22 points in each of the four main topic areas of the course. There were enough activities available in each topic area to complete much more than 22 points, however, and this is where the evaluation system for the course becomes quite unique.

By aiming for more than 22 points on a particular assignment, students could essentially do ‘extra credit’ work, creating a buffer that would allow for the possibility that they might fall short of their expectation in one activity (for example, earning only a “Basic’ rating when they had hoped for an “Exemplary”), yet still, by virtue of having targeted extra points, wind up with full marks on the assignment overall.

Naturally, this took a fair amount of careful explanation, and some students initially did find it confusing, but we were able to refine our instructions and explanations of the system well enough so that now, as the course enters its third year of operation, we get very few requests from new students seeking help understanding the system.

With 4 major assignments, each worth a maximum of 22 point, we had distributed 88% of the course grade weight. The remaining 12% I had set aside for something else I wanted to include.

One of the things which was both a benefit and a limitation of the course design, was that students would see only what they chose to see, so I wanted to enforce a degree of interaction between the students and also facilitate them seeing what others were doing in the course.

For the final project, students were therefore required to choose several of their own completed activities, write a brief statement of the learning objective the material would attempt to achieve, and post them in a special conference
we created in our learning management system (Moodle) for this purpose. These “Learning Objects” were then available for all other students in the course to view.

The second part of this final assignment was to choose several items from among these posted ‘Learning Objects’ and review them, which meant reading, watching or listening to the submitted item, depending on its format, and writing a review. There were some general parameters defined for the format and content of the reviews, but again, students were free to choose what they themselves posted from their own work, and also what they chose to review. The stated learning objectives for their own posts, plus their reviews of other students’ posts, were combined into a single portfolio for submission.

Given that the textbook for this course was, in a sense, the entire Internet, it was obvious from the start that there could be no exams or quizzes, for how can you make a fair exam question in a course where every student is learning a different specific body of content?

Presenting this course design to my colleagues was an interesting experience.

Here was a course with no text, no exams or quizzes, and no common body of content except in a very general sense, no uniform format for assignment submission and an evaluation system that allowed students to earn more than 100% on individual assignments (of course, marks over 22/22 on individual assignments were not actually awarded, but, as previously mentioned this allowed students to lose marks on a specific activity and still wind up with a perfect grade on the overall assignment). To their credit, and my pleasant surprise, the other members of the Faculty were generally supportive, and the course was allowed to proceed to production as designed.

I also had some trepidation about the reaction of students, though I was fairly confident the course would be well received once students really understood what we were trying to do. Since the course opened I would have to say that it has received the most positive feedback of any course I have developed or managed. It also produces the highest grades, partly by virtue of the ‘extra credit’ aspect of the evaluation system, but also because, when faced with an open ended, rather loosely defined task, and especially one of their own choosing, students simply work harder, go further, and deliver more. They err on the side of caution, wishing to ensure that they achieve that “Exemplary” grade, but also, they are actually interested in what they are doing because they got to choose the specific focus of their activity.

Here is a selection of comments received from students as they completed the course:

“I just wanted to take the opportunity to let you know that at the beginning of the course I was quite apprehensive because of the course format. That feeling was due to the difference of other courses I have taken. I just wanted to let you know that I really enjoyed the course and the learning opportunities are endless with this format. I have learned quite a bit about HCI and not just the topics I chose for assignments. Through poking around and researching, I learned much more than any text book could have provided.”

“I just wanted to say I enjoyed the course and thank you for this type of learning opportunity.”

“This has been an ongoing "quest" since I first started the course and quite frankly Facebook groups, Pinterest and library research proved the most useful activities. Despite the forum postings on the landing that received little response I found other "social" groups and discussion areas quite satisfactory for navigating to credible resources. I often speak with several colleagues quite intimately about computer technology and those conversations are indispensable as well. Thank you for the opportunity to pursue some interesting and useful research related to HCI.”

“just wanted you to know I'm enjoying this class and have worked hard on various assignments!”

“I'm glad that you enjoyed my assignments; I had a lot of fun as well - a lot more entertaining than the operating system work and calculus I've got this semester. With luck I'll get another one of your courses, though. Thanks again!”
“I really enjoyed the course, it's well put together and I liked the hands on approach for MIT Scratch and Logo.”

“It was a rather unconventional and unusual course, but I don't think it could be done any other way.”

References


Appendix: Interactive technologies course details

Below are the topical outline and some examples of the specific activities prescribed in the course Interactive Technologies.

Outline of main topic areas

**Topic 1 - The context of human-computer interaction**
- History and Development of Computer Interfaces
- The Next Generation of Interfaces
- Ergonomics

**Topic 2 – Input and output**
- Input at the Programming Level
- Input Devices and Data Types- From Typing to Sensing
- Output – Not Just Information Anymore

**Topic 3 – Using and evaluating application interfaces**
- Windows and other desktop interfaces
- 2D and 3D Interfaces

**Topic 4 – Creating user interfaces**
- Principles of Sound Interface Design
- Creating Simple HTML Web Interfaces
- Image Maps
- Using Simple Programming Tools to Design Interactive Applications or Interfaces

Course activities

The activities in the course each have specific Learning Objectives, and specific deliverables, which may be in the form of illustrated articles, PPT presentations, webpages or videos. Two of the activities have computer programs as their products, created using very simple programming interfaces.

Many activities require that the student simply research and compile information on certain specific aspects of the course Topics, others require the student to install and work with specific software tools such as gaming, programming or multimedia applications.

Activities are organized by Topic and general subtopic where appropriate. Students are required to complete a minimum number of 3 activities from each of the 4 course topics, however they may choose which specific activities in each Topic area they wish to complete.

**Topic 1 – The context of human-computer interaction**

**History and development of computer interfaces**
- 1.1 – Create a timeline of computer interface technology
- 1.2 - Profile an HCI developer
- 1.3 - Man versus technology

**The Next Generation of Interfaces**
- 1.4 - Develop a presentation on emerging HCI technology
- 1.5 - Describe an ‘ultimate level’ of human computer interaction

**Ergonomics**
- 1.6 - Research and document an ergonomic issue in HCI
Topic 2 - Input and output

**Input at the programming level**
2.1 - Research and report on rapid interface design tools

**Input devices and data types - from typing to sensing**
2.2 - Install and test a freeware voice command/dictation program
2.3 - Describe interaction methods for a non-traditional computing device

**Output – Not just information anymore**
2.4 - Describe a specific computer managed manufacturing process
2.5 - Describe a large scale display system

Topic 3 - Using and evaluating application interfaces

**Windows and other desktop interfaces**
3.1 - Describe graphical user interfaces – windows
3.2 - Customize and adapt the windows interface

**2D and 3D Interfaces**
3.3 - Describe a specialized gaming interface
3.4 - Evaluate a VR interface – Google Earth
3.5 - Evaluate a VR interface – Second Life

Topic 4 - Creating user interfaces

**Principles of sound interface design**
4.1 - Review an article or lecture on human-computer interaction

**Web interfaces**
4.2 - Create a web interface for course topic explorations
4.3 - Create a web interface for image display

**Image maps**
4.4 - Create an intuitive image map interface

**Using simple programming tools to design interactive applications or interfaces**
4.5 - Learn and use a programming interface – LOGO
4.6 - Create an interactive application using MIT Scratch

**Evaluation of activities**

Each activity allows for a range of points, depending on the difficulty of the activity and the rating assigned by the marker. For example, simple activities may earn from 1-3 points, moderately difficult activities 3-5 point and difficult activities 4-7 points.

Only a maximum of 22 points may be scored on a given assignment, but students are free to undertake activities for more points than they can score.

Evaluation is based on the difficulty level of the activities completed, as well as on the standard achieved in the completion of the submitted activities. Each submitted activity is rated as meeting the Basic, Complete, or Exemplary standard and assigned an appropriate number of points accordingly. These standards are described below:

- **Basic Standard:** The work demonstrates a minimum response to the requirements for the activity. It includes the specified elements, but shows little or no creativity, design, or media use; minimal detail; and/or incomplete citation of sources.
- **Complete Standard:** The work demonstrates a thorough response to the requirements for the activity; some effort in design and media use; thorough detail; full citations; and effective organization.
- **Exemplary Standard:** The work demonstrates an outstanding response to the requirements of the activity. It is well organized and detailed; shows insight as well as knowledge; is creative in design and media use; and has an original or challenging topic.

Note: Activity submissions that fall below the Basic standard are returned ungraded to the student with suggestions for improvement. Students may resubmit each activity only once, after which it will be graded.
Collaborative mLearning: A Module for Learning Secondary School Science

Dorothy DeWitt*, Saedah Siraj and Norlidah Alias
Faculty of Education, University of Malaya 50603 Kuala Lumpur, Malaysia // dorothy@um.edu.my // saedah@um.edu.my // drnorlidah@um.edu.my

*Corresponding author

ABSTRACT
Collaborative learning has been shown to be effective for the construction of knowledge. In science instruction, collaboration and knowledge-construction needs to be done in the language of science. A collaborative mLearning (CmL) science module employed uses three computer-mediated communication (CMC) tools: wiki, discussion forum and text messaging. This study seeks to determine the forms of communication and learning in the use of these CMC tools in the CmL module. Twenty (20) Form 2 students of different science abilities participated in the study. Data were collected from student interviews; online communications on the wiki, discussion forums, and text messages; students and researchers’ journal records; and a survey of students’ perception of communication with the CMC tools and learning. The findings showed the learners’ frequency of communication was highest in the wiki and text messaging. The combination of three CMC tools was effective as it catered to learners’ preferred learning styles. Group work and the collaborative activities enabled learning. The CmL module was effective for learning as verified by the improvement in post-test results. The findings of this study provide insights into group interactions in a CmL environment and show that peer interactions scaffold learners in building their knowledge in science.

Keywords
Collaborative mobile learning, Computer-mediated communication tools, Mobile learning, Informal learning

Introduction
Collaborative learning enables learning experiences to be interpreted for the construction of knowledge (Palloff & Pratt, 1999). However, the effect of computer-mediated communication (CMC) tools for collaborative learning, or collaborative mobile learning (CmL), is less explored. Studies have shown that CmL is useful for peer support in scaffolding learning (Bottick, Looi, & Wong, 2011; Timmis, 2012), generating ideas (So, Tan, & Tay, 2012), and knowledge-creation (Rogers, Connelly, Hazlewood, & Tedesco, 2010). Different CMC tools have different affordances: discussion forums (Guzdial & Turns, 2000; Slotta & Linn, 2000), wikis (Bonk, Lee, Kim, & Lin, 2009; Pifarré & Li, 2012; Zhang, Scardamalia, Lamon, Messina, & Reeve, 2007) and text messaging (Capuano, Gaeta, Miranda, & Pappacena, 2005; Timmis, 2012) have been used for learning.

CMC tools have been used for teaching science (Guzdial & Turns, 2000; Slotta & Linn, 2000). However, text messaging, and mobile devices which have been for used for learning language (Arrigo, Gentile, Taibi, Chiappone, & Tegolo, 2005; Botticki, Looi, & Wong, 2011; Capuano et al., 2005; Gerosa, Filippo, Pimentel, Fuks, & Lucena, 2010), are not used much in science instruction. A combination of two tools: text messaging with a wiki (Arrigo et al., 2004), and text messaging with a discussion forum (Gerosa, Filippo, Pimentel, Fuks, & Lucena, 2010; Rau, Gao, & Wu, 2008), has been shown to be effective for science learning. This study will investigate the combination of three tools in developing the CmL Science module.

Collaborative learning is rarely implemented in the Malaysian scenario. Teachers perceive that there is insufficient time to complete the science syllabus and allot little time for social interaction in the science classroom. Teachers emphasize the memorization of facts rather than the scientific processes (Chong, 2005). Hence this study seeks to extend previous research by investigating the use of three CMC tools for learning in the CmL environment: the wiki, discussion forum and text messaging. In addition, it will determine if CMC tools are effective for collaborative and mobile learning in science, and whether these interactions can take place out of the formal classroom environment.

Collaborative learning is the acquisition of knowledge, skills and attitudes as a result of group interactions (Johnson & Johnson, 2004). When CMC tools are employed for interactions, learning becomes mobile and hence, collaborative mobile learning (CmL) occurs; CmL allows group interactions outside the formal classroom
environment and learning happens anytime, anywhere (Ally, 2004; Siraj, 2005; Siraj & Alias, 2005) and is situated in the environment (Chang, 2010; Jeng, Wu, Huang, Tan, & Yang 2010). This study is significant as it would determine whether CmL can be conducted in Malaysian schools.

The language of science

In scientific discoveries, scientists collaborate with other scientists through scientific processes (Emdin, 2010; Nielsen, 2012; Sharma & Anderson 2009). Hence, the language of science is required for discussions and collaboration.

A similar approach of discovery and collaboration should be used for science instruction in schools. Instruction should be focused on scientific methods and processes, built upon social interactions (Hogan & Fisherkeller, 2005; Sharma & Anderson, 2009). Hence, science learners need to be able to communicate socially, interact and debate issues regarding science and society, as well as use science for their personal needs. Social interactions enable learners to attempt to link the newly acquired knowledge with their existing knowledge, and be scaffolded individually (DeWitt & Siraj, 2008).

The language of science enables scientists to construct science concepts. Scientific verbal knowledge is required for planning and for sharing ideas (Ellerton, 2003; Hoyle & Stone, 2000). While scientific terms can be defined and taught formally in the classroom, the vocabulary and language structures in science, which enable critical thinking, are acquired informally. The advantage of student-centered discussions is that learners construct meaningful phrases and sentences to communicate, and resolve differences of opinions to reach mutual understandings as science concepts are developed (Hoyle & Stone, 2000; Karpov & Haywood, 1998).

The vocabulary and structures of the language of science are informally modeled by learners through interactions with the materials, teachers and peers (Karpov & Haywood, 1998). Vygotsky’s view is that scientific knowledge and procedures should not be taught directly but should be constructed by learners in the course of a discussion (Karpov & Haywood, 1998). The interactions between other learners, the tutor and learning materials on a suitable platform can enhance the learners’ current understandings of concepts and principles.

Design of instruction for learning science

The CmL Science module is based on social constructivist learning theories. Sufficient activities are provided for patterning and modeling the language of science, with individualized support and scaffolding, as well as opportunities for discussion to assist learners in building their personal understanding of scientific concepts and principles (Ellerton, 2003; Hoyle & Stone, 2000). The activities allow learners to link the science knowledge with their own personal experiences (Ellerton, 2003). These social interactions motivate and engage learners in carrying out activities while building meaningful science knowledge (Brown, 2006).

CMC tools have been used for teaching science. The Knowledge Integration Environment (KIE), a platform for web resources, has a discussion forum for the social environment and context for collaborative mLearning in science, has been shown to be useful for learning science (Slotta & Linn, 2000). CaMILE, another platform for collaborative mLearning which uses discussion forums for learners to collaborate on science inquiry projects, is effective for learning (Guzdial & Turns, 2000). However, both these platforms do not have text messaging.

Text messages may be combined with other CMC tools in a science module. Presently, text messaging has been used for language instruction where messages can be pushed as textual learning objects (Capuano et al., 2005). Text messaging has been combined with wikis (Arrigo et al., 2005), and with discussion forums (Rau, Gao, & Wu, 2008). The use of text-messaging motivated learners, and with online discussion groups, improved examination performance (Rau et al., 2008). Hence, there is a possibility that a combination of CMC tools, with text messaging, could be used for learning science.
Research is lacking on the use of a combination of CMC tools on a CmL platform, especially in the Malaysian context. It is hoped that this study will provide insights in the use of a combination of CMC tools, namely wiki, discussion forums and text messaging.

**Theoretical framework**

*Collaborative learning*

Collaborative learning is dependent on social interactions. Collaborating in groups has been shown to improve memory, produce fewer errors and motivate learners (Bligh, 2000). Background factors, such as age, activeness and values; internal influences such as leadership and communications; and consequences on why collaboration is required will influence the group interactions (Tubbs, 1995).

Communications is not just the transfer of information but interactions that enable the process of meaning-making in science (Sharma & Anderson, 2009; Tubbs, 1995). As learners interact, both face-to-face and online, and reflect on their discussions, a learning community for sharing learning experiences is built (So & Bonk, 2010, Palloff & Pratt, 1999). CMC tools facilitate CmL outside the classroom (Anastopolou, Sharples, Ainsworth, Crook, O’Malley & Wright, 2012; Arrigo et al., 2005; Capuano et al., 2005; Guzdial & Turns, 2000; Slotta & Linn, 2000).

New models of CmL are required as CMC may not be supported by conventional face-to-face learning environments (Li, Dong & Huang, 2011). The challenge is addressing the diversity of online learners to achieve a common goal (Kuo, Huang, Chen, & Chen, 2102; Palloff & Pratt, 1999). Similar to conventional learning, learners need to conform to group norms according to standard behaviors and conduct, to avoid conflict and tension in the group (Tubbs, 1995). Adapting to the use of new technologies may cause disorientating dilemmas and psychic distortions (Palloff & Pratt, 1999) which requires learners to change their attitudes to learning (Tambouris, Panopoulou, Tarabanis, Ryberg, Buus, Peristeras, Lee, & Porwol, 2012). The instructor’s role is to support and allow a conducive practicing environment (Mostmans, Vleugels, & Bannier, 2012).

Support in a CmL environment is required to engage learners outside of the social context of the classroom (Li et al., 2011). Hence, a structured support system is required for scaffolding learning, supported by peers (Boticki, Looi, & Wong, 2011; Timmis, 2012).

In social constructivist theory, dialogue and interaction internalizes learning (Gredler, 1997; Schunk, 2000). Cultural tools such as computers and mobile phones; and abstract social tools, such as language, assist in developing the learners’ thinking. CMC tools enable cognitive change in the learner as ideas are exchanged and debated upon to create new knowledge (Zhu, 2012; Gredler, 1997; So & Bonk, 2010; So, Tan, & Tay, 2012; Rogers, Connelly, Hazlewood, & Tedesco, 2010). Although CmL enhances student learning achievement, it may be influenced by the culture of the community (Zhu, 2012). It was noted that there might be gaps between western and certain eastern cultures (Zhu, 2012).

*The social constructivist theory of learning*

Social dialogue and interaction internalizes learning (Gredler, 1997; Schunk, 2000). In the external environment, cultural tools such as computers and mobile phones; and abstract social tools, such as language and the school, assist in developing the learners’ thinking. CMC tools encourage the process of cognitive change in learners (Gredler, 1997).

In this study, learners interact socially in a group through peer collaboration (Schunk, 2000) to complete tasks which develop their understanding. Instructional scaffolding related to science is given for support to accomplish the tasks. This is in line with collaborative learning which assumes that human intelligence originates in society or culture, and individuals’ cognition results from interpersonal interaction within that culture and community (Vygotsky, 1981).
The research questions

The purpose of this research is to determine the forms of communication and learning among learners during the implementation of the CmL Science module. This study seeks to answer the following research questions:

- How are the learners communicating in the CmL Science module?
- How do learners learn with the CmL Science module?

Methodology

Design of the study

The study is part of developmental research where the CmL module was designed (Wang & Hanafin, 2005). An urban secondary school with a multiracial composition was selected for implementation of the CmL Science module in the topic of nutrition. A survey of the communication tools the learners accessed contributed to the design of the CmL module. The module was implemented with twenty Form 2 students. The participants’ use of the module and their perceptions regarding learning and communications were captured through observations, online communications, interviews and a survey of the module. The responses given were triangulated with data from online communications in the module and a pre-test and post-test design.

Data collection

A pre-test on the knowledge items in the module was conducted before the module implementation. During implementation, data from the online communications on forums, wikis, and text messages; as well as participants’ and researchers’ journals, were recorded. At the end of the module, a post-test was conducted followed by a survey to determine the learners’ perception of their understanding and support required for learning. The participants were then interviewed to glean further information. Triangulation of data was done through analysis of online communications to verify that the users were learning.

Development of the module

The CmL Science module took into account the CMC tools the learners accessed. This module was on a webpage, with links to resources including content, videos, animations, and tasks on CMC tools. The topic of Nutrition was selected as learners had misconceptions in this topic. Secondary school children were confused about the concept of food: Water and vitamins were inaccurately considered as food (Lee & Diong, 1999), and many had bad dietary habits (Melby, Femea, & Siacca, 1986). Even trainee teachers had misconceptions on the function of proteins (Lakin, 2004). In addition, Nutrition was rated as the most difficult topic in science since it involved a lot of factual knowledge (DeWitt & Siraj, 2007).
CMC tools in the CmL science module

The CmL Science module consists of online lessons and face-to-face meetings (Table 1). Activities in the CmL module are the problem task, assigned as group work on the wiki; related questions on the discussion forum; and text messaging quizzes pushed to individual learners’ mobile phones (Figure 1). Feedback and support in all the tasks were given by the tutor. Table 2 provides a summary of the CMC tools used in the CmL Science module.

Table 1. List of lessons in the CmL science module

<table>
<thead>
<tr>
<th>Lessons</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>Orientation to module</td>
</tr>
<tr>
<td>1</td>
<td>The classes of food</td>
</tr>
<tr>
<td>2</td>
<td>Special diets (balanced meals)</td>
</tr>
<tr>
<td>3</td>
<td>Tests for food classes</td>
</tr>
<tr>
<td>4</td>
<td>Counting calories</td>
</tr>
<tr>
<td>5</td>
<td>Food in customs and cultures</td>
</tr>
<tr>
<td>Final</td>
<td>Summary</td>
</tr>
</tbody>
</table>

*Face to face meetings

Table 2. Tools and activities in the CmL science module

<table>
<thead>
<tr>
<th>Platform</th>
<th>CMC tool</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>The CmL Science Module</td>
<td>Wiki</td>
<td>Group Activity: Groups of 3 – 4 students.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Task is to analyze the nutrients in a given meal and post findings on the wiki. Each group has a wiki page for the task, and can add additional pages.</td>
</tr>
<tr>
<td></td>
<td>Discussion forums</td>
<td>Individual Activity: Discussion questions based on the lessons.</td>
</tr>
<tr>
<td></td>
<td>Text messaging</td>
<td>Individual Activity: SMS Quiz Questions pushed to students with immediate feedback.</td>
</tr>
</tbody>
</table>

Module implementation

The CmL Science module was implemented with a group of twenty Form 2 students from a selected urban school with a multi-racial composition. The participants were volunteers with equal numbers of high, medium and low-achievers in science.

Most of the students in the selected school (81.6%) owned a mobile phone, and more than half had access to computers (63.9%). Laptops and mobile phones were made available during the implementation for students to access. Mobile phones were required for text messaging after school hours. Participants who needed a mobile phone could borrow one.

Data analysis

Data collected from the interviews and interactions in the CMC tools, participants’ and researchers’ journal records were transcribed and analyzed to identify how students in the context of the study communicated and learnt in the CmL environment. In addition, the results of the pre-test conducted before the implementation of the module, was compared to the post-test at the end of the module to determine if there was a difference in the scores. This would determine if learning was effective. The survey conducted after the implementation of the module, was analyzed using descriptive analysis.
Results and discussion

The discussion is focused on studying two areas in CmL module: communications and learning. The findings may provide insights regarding the communications in the CmL module and the contribution to learning science.

Communications in the CmL module

In order to investigate how learners communicate in the CmL module, two domains were analyzed: the frequency of use of the CMC tools, and types of group activities.

CMC Tools

The frequency of using the CMC tools for the assigned tasks varied among participants (Figure 2). Participants were most active in the SMS Quiz (65 log-ins). The increased participation in the SMS Quiz was due to the accessibility of the mobile device: “Our phone is just in our pockets, so we can reply immediately. Using the computer to access the internet takes time”. In addition, personalized feedback was given during the SMS Quiz.

There were a large number of postings on the discussion forum (33 responses) but analysis showed that only an average of 8 participants had responded. Most of responses to the questions were shared responses posted by two or more learners, indicating the participants preferred group work. Participants collaborated with their peers before posting in the discussion forum.

The interviews revealed that the lack of activity on the forum compared to the SMS Quiz was because the forum questions were difficult, and the reluctance of participants contributing after reading others’ posts:

S: Because I don’t really know the answer.
DD: But do you read your friend’s answers?
S: No. It’s because if I read their answers, they may think I’m copying their work.

The concept of copying answers was a situational cue from prior experience in the classroom. Learners lacked confidence and had the fear of being accused of cheating.

DD: Did you read the other participants’ answers?
I: Sometimes. It’s alright to read if I don’t know the answer. But if I do know the answer, then I’ll just write it, and won’t read the others’ work.

The wiki was accessed less often (17 log-ins) by only 12 participants compared to the other tools. However, the participants were engaged with the task as in the first week, they had started to edit their group wiki, changing font types, adding colors, graphics and animation. Only in one group, the members had difficulty cooperating and no activity was detected.

In summary, the mobile phone was a personal device which was easily accessible, hence most frequently used. There were fewer posts in both discussion forums and wiki as the learners were working in groups. However, both these tools were useful for group responses. In addition, the wiki enabled artifacts such as graphics and text to be used for communication.
**Working in groups**

Participants preferred working and collaborating in groups (Figure 3). The group task on the wiki was attempted by most of the participants. In addition, contributions to the discussion forum were posted by groups of twos or threes participants.

The frequency of participation in the wiki differed among the groups. In two groups (Groups 1 and 5), all members participated online while in other groups (Groups 1 and 4) only one member was assigned to post the solution after the face-to-face discussion. The reason for the lack of online activity in some wikis is because one member is assigned to the posting, as explained: “We do the discussions in a group. We don’t use the computer yet. We talk and sit face-to-face, and one person will write what we discussed about.”

Group discussions and collaboration was also conducted through other CMC tools. Group 1 used Instant Messaging for discussions. “Sometimes I’m online at MSN, then I ask my friends what I should do. And my group members tell me what to do.”

There was a preference for group work. Results from the survey showed participants liked group work for both the online tasks (80%) and discussion forum (87%), and did not perceive group work as difficult (67%) (Figure 3). Although there were some who did not like to depend on others (“I dislike the online tasks because I need the group members”), most perceived group members as willing to contribute ideas (87%). Group discussions were conducted in face-to-face discussions and extended to the text messages on SMS and on MSN chat.

**Learning with the CmL science module**

Learning with the CmL Science module is investigated under two domains: Understanding science and support for learning.

**Understanding science**

The CmL Science module improved understanding as shown in the results of a survey, interview as well as pre and post-test results of students.

The survey showed most learners (84.3%) believed their understanding in science improved on completion of the module (Figure 4): “Because when I do this module, it improves my knowledge. In addition, the module helped me to revise the topic.”
Further, in using the module, learners interacted with online learning materials and books. As one learner reported: “It makes me open my book: I won’t open books if I don’t have exams.” In addition, they viewed other participants’ answers in the discussion forum.

![Figure 4. Learning using the collaborative mLearning science module](chart)

The learners believed that the discussions in the CmL Science module facilitated learning. “Honestly, I think the SMS Quiz makes me remember. Because we like to talk to our friends, and this chit-chat helps us recall better. Well, one thing for sure, with the questions, you can always ask people, or you can refer to your Science text book. At least it helps us to brainstorm a bit.” One participant explained how the module had helped him in answering test questions: “There’s this one question on the walls of the organ in the objective section. I picked C as the answer. And in the subjective questions, they asked about the name of the movement of food in the organs, so I recalled it- peristalsis.”

A pre-test and post-test on the similar concepts showed an increase in the mean scores of participants. The $t$-test for independent samples for statistical analysis was not computed as only 16 participants completed both pre-test and post-tests. However, the difference in the mean scores showed an increase which might indicate the module was effective for learning science concepts (Table 3).

<table>
<thead>
<tr>
<th></th>
<th>Mean scores</th>
<th>Increase in mean scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre test</td>
<td>61.97</td>
<td></td>
</tr>
<tr>
<td>Post test</td>
<td>83.07</td>
<td>21.09</td>
</tr>
</tbody>
</table>

Note: Number of participants, $n = 20$

The CmL module enhanced understanding of science concepts related to nutrition. Learners made references to materials, both online and print; and had discussions in science. Reference to materials enabled learners to formally model their answers while informal discussions in the search for knowledge contributed to understanding science. In this way, learners are able to build knowledge and scientific verbal language.

**Support for learning**

Support during the learning process provided the scaffolding for the novice learner to reduce the zone of proximal development (Schunk, 2000). The expert might be their peers or the instructor from whom they receive support and verification for what they were doing. The survey and interview would be used to determine if learners required support while additional data would be collected through observation of the online interactions to determine learners’ interactions.

Learning is supported through their peers’ interactions as evidenced from the group tasks on the wiki. Group work seemed to be effective for learning (Bligh, 2000; Sharma & Anderson, 2009; Tubbs, 1995). Group members who were cooperative were able to collaborate well and contribute to better solutions on the wiki. In addition, groups with members of similar interests and background were more active and seemed to perform better. This was because group members who worked well together online had a shared goal.
In the CmL module, the tutor monitored the learners’ activities online and provided support when necessary. Assistance was provided when asked, or when the tutor perceived the learner required assistance. Scaffolding in the form of exemplars, hints and suggestions of strategies were given to assist in constructing the solution. In the discussion questions, the tutor suggested aspects of the question which the participant might not have considered.

However, learners perceived they required more support from the tutor. Novice learners wanted more help and feedback from the tutor on the wiki (73%) and the discussion forum (80%) (Figure 5). The tutor was the “provider of knowledge” and learners expected the tutor’s continuous presence. In addition, Malaysian learners seem to require more scaffolding, perhaps because of the social norms, beliefs, and behavior of a teacher-centered education system, where the teacher is the source of knowledge.

Scaffolding needs to be gradually withdrawn to allow learners to build their knowledge through the thinking processes. Learners were not used to the absence of a teacher. However, learners will be forced to think for themselves without the teacher’s continuous presence. As one learner reflected: “When the teacher’s not there, I have to go solo and have to think to get the answer.”

Scaffolding supports the informal learning of science through conversations, discussions and other online communication as it provides opportunities for modeling science in the learning materials and the discussions in the CmL Science environment. These patterns of language and concepts formed when learners collaborate on their tasks contribute to the understanding and learning of science.

**Implications and conclusions**

The findings of this study indicate that the CmL Science module can be implemented in secondary school science. Learners participated in the discussions in science using CMC tools showing that CmL can effectively be used to support communication for learning (Anastopolou et al., 2011; Arrigo et al., 2005; Guzdial & Turns, 2000). Communications on the wiki and discussion forums could be viewed by the public and formed a permanent representation of the tasks, or the answer to the problem. At the same time, informal communications in constructing the solution contributed to the thinking process. These interactions and discussions were beneficial for learning and constructing science knowledge (Hoyle & Stone, 2000). This was proven from the results of the pre-test and post-test. Although there were only twenty participants, the findings are relevant as the participants are of different science abilities. However, future studies could be done to determine the effectiveness of the CmL module with a larger sample.

Private discussions were also conducted, both face-to-face and through text messages. The discussions for solving the task enabled the learners to have a shared goal and form a learning community which contributed to knowledge building (Johnson & Johnson, 2004; Kuo, Hwang, Chen, & Chen, 2012; Palloff & Pratt, 1999).
The comparison of the frequency of use of the CMC tools showed text messaging was the preferred communication tool. Text messages were pushed to learners who owned the cultural tool, the mobile phone, which was with them most of the time (Capuano et al., 2005). Hence, the ease of accessibility and use was a factor for increased use. On the other hand, the wiki and discussion forum was a little more difficult to access as a computer was required. The wiki was preferred compared to discussion forum as it was easier, more attractive to use, and encouraged group work.

Previous studies have combined the use of several tools for learning. The use of a combination of three tools was more effective than one or two. Not all learners used all the CMC tools during implementation. Hence, by providing a choice of several tools, we allowed learners to respond using the tool most suitable or preferred for learning. Wiki lends itself to collaborative problem solving; discussion forums were best for debates and arguments while text messaging was for factual and conceptual knowledge. This study also confirms that text-messaging could be used for informal discussions for building knowledge, as well as a stimulus through formal learning when quizzes are answered.

The preference for different CMC tools was related to the learners’ learning styles and convenience. The influence of learning styles and the convenience of the tool could be an area for further research. In addition, further studies could be done to investigate if there are any significant differences when only one, two or three CMC tools are used.

Use of CMC tools affords collaborative mobile learning (Guzdial & Turns, 2000; Slotta & Linn, 2000). Learners prefer working in groups when attempting the tasks with the CMC tools. Collaborative mlearning, as a result of using CMC tools for group interactions, was effective (Ally, 2004; Johnson & Johnson, 2004; Siraj, 2005; Siraj & Alias, 2005). There was some disorientating dilemmas in some groups when the group did not cooperate and coordinate well (Tubbs, 1995). This implies that group dynamics and group building should be conducted in order to establish specific behavioral norms. Support from the instructor could maintain the group’s dynamics.

In addition, silent observers, who did not seem to participate but observed the social interactions, were also involved in the informal learning process. This was evidenced as when they viewed other participants’ answers, the patterns and use of scientific verbal language was observed (Karpov & Haywood, 1998). It was also noted that lack of participation in discussions could be attributed to the learners’ perception that there was only one correct answer in science. These learners have to be given more scaffolding and encouragement to participate in the communication and learning process and to be made aware of the nature of science knowledge (Schunk, 2000). In addition, considerations may have to be made in the social and cultural tools to include more exemplars and guidance for discussion questions (Gredler, 1997). Orientation to the module should also include more group activities to get to know the members and develop a shared goal in learning (Tubbs, 1995). Future studies should investigate the difference in having heterogeneous groups, and groups which are formed according to participants’ choice in CmL. In addition, methods to engage the silent observers in order to ensure full participation have to be considered.

The findings of the study reinforce the fact that communicating in science is important for learners to plan, share ideas, develop their understanding and promote critical thinking through the language of science in a collaborative environment (Ellerton, 2003; Hoyle & Stone, 2000). The communication contributed to the learning of science. Hence, the CmL module can be used for learning science out of school hours, to address the problem of lack of time for classroom discussion.

The CmL module was developed for online communication and collaboration with appropriate tools to encourage learning of science. Most learners seemed to prefer to collaborate while attempting the activities. Learners discussed when solving the problem task on the wiki and attempted answering the discussion questions in small groups. The patterning and modeling of the language of science was developed during group discussions (Karpov & Haywood, 1998).

In summary, social learning in the use of the collaborative mLearning module could be used to teach science to address the learning needs in the field. This aspect of social interaction for building knowledge thorough formal and informal learning can be extended to other subjects as well. Elements in the environment, including participants’ answers were artifacts to “mediate” learning. Hence, in the CmL Science module, non-participation did not mean the learner was inactive. Learning could take place formally and informally when other learners’ answers and interactions mediated learning.
Acknowledgments

This work was supported by the University Malaya Research Grant RG386-12HNE and Flagship Research Project FL019-2012.

References


Facilitating English-Language Reading Performance by a Digital Reading Annotation System with Self-Regulated Learning Mechanisms

Chih-Ming Chen1*, Jung-Ying Wang2 and Yen-Chang Chen3

1Graduate Institute of Library, Information and Archival Studies, National Chengchi University, Taipei City, Taiwan, ROC // 2Department of Multimedia and Game Science, Lunghua University of Science and Technology, Taoyuan, Taiwan, ROC // 3E-learning Master Program of Library and Information Studies, National Chengchi University, Taipei City, Taiwan, ROC // chencm@nccu.edu.tw // wyy@mail.lhu.edu.tw // yenchang65@yahoo.com.tw

*Corresponding author

(Submitted October 02, 2012; Revised December 29, 2012; Accepted March 31, 2013)

ABSTRACT

Since English has been an international language, how to enhance English levels of people by useful computer assisted learning forms or tools is a critical issue in non-English speaking countries. Past studies confirmed that reading articles with annotated contents enable knowledge sharing and improve the reading comprehension of learners. However, the self-regulated learning (SRL) ability of individual learners when reading learning materials and contributing annotations becomes a key factor affecting reading performance. Thus, this work proposes a SRL mechanism combined with a digital reading annotation system (DRAS) to enhance Grade 7 students to generate rich and high-quality annotations for promoting English-language reading performance. To evaluate the effectiveness of the proposed system, this work adopts a quasi-experimental design to assess an experimental group and control group learners who respectively use the proposed DRAS with and without the SRL mechanisms when reading English-language texts online. Compared with the control group learners, experimental results demonstrate that the reading comprehension and reading annotation abilities of the experimental group learners were significantly improved. Analytical results also confirm that gender differences in reading comprehension and annotation ability existed when using the proposed DRAS with and without the SRL mechanisms to read English-language texts online. Experimental results also show that significant differences existed in the reading comprehension and annotation abilities of learners with good and poor SRL abilities in the experimental group. Additionally, the reading annotation ability of learners in the experimental group was significantly correlated with reading comprehension.

Keywords

Self-regulated learning, Digital reading, Reading annotation, Second language learning

Introduction

As English is now considered the dominant business language worldwide, how to enhance the English-language reading ability of students is of utmost importance, as good English-language reading ability improves an individual’s competitive advantage (Mahapatra et al., 2010; Risko et al., 2011). To enhance English reading performance, students typically use annotation techniques, such as underlining, highlighting, notes, and summarizing, to support their reading in traditional printed books. These annotation techniques are very helpful in understanding and memorizing the reading contents in an article (Hoff, Wehling & Rothkugel, 2009). In recent years, many computer-assisted reading annotation systems have been developed to assist learners in learning by reading digital texts (Belz, 2004; Patrick Rau, Chen & Chin, 2004; Mendenhall & Johnson, 2010; Johnson, Archibald & Tenenbaum, 2010; Wolfe, 2008; Johnson & Nádas, 2009) because using computer monitors and other digital reading devices for reading or browsing texts has gradually become common reading modes. Hence, this work developed a novel web-based digital reading annotation system (DRAS) for individual learners to enhance their reading comprehension of English-language texts. The proposed system designs several annotation functionalities that can support individual or collaborative annotation of digital texts and saves annotations in a log database. To promote reading performance, learners can utilize the Internet anytime and anywhere to share their annotations and interact with others. That is, the proposed web-based digital reading annotation system provides a flexible learning environment that improves reading comprehension and performance of individual learners.

However, when utilizing the proposed DRAS to promote English reading performance, learners must often perform self-directed learning. Therefore, the SRL abilities of individual learners associated with actively learning the reading materials and contributing annotations are main factors that affect learning performance. Many studies have demonstrated that learners who are unable to self-regulate their learning strategies tend to misunderstand complex
topics (Hannafin & Land, 1997; Jacobson & Archodidou, 2000). Thus, this work proposes a SRL mechanism combined with DRAS to enhance English-language reading performance. The research questions of this study include whether the proposed DRAS with SRL mechanisms can promote reading comprehension and annotation abilities of individual learners, whether gender difference and correlation between reading comprehension and reading annotation ability exist, and how different SRL abilities affect reading comprehension and annotation abilities of individual learners.

**Literature review**

**Web-based language learning**

In recent years, conventional computer-assisted language learning (CALL) has gradually moved toward web-based language learning (WBLL) because WBLL provides language teachers with network-based teaching environments in which they can assign meaningful tasks and use various materials for language learning (Son, 2008). Particularly, the hypermedia character of the Internet has markedly enhanced the power of CALL by allowing learners to explore and discover their learning processes and offering learners access to online resources (Son, 2008). Additionally, WBLL provides learners with an interface for interaction and gives students and teachers alternative ways to communicate. Khan (1997) indicated that web-based instruction helps learners complete a series of instructional activities, and helps learners increase the number of opportunities for constructing and sharing their knowledge with others. In other words, as current WBLL paradigms offer advantages in promoting language learning effectiveness and teaching, their impacts on language learning should be investigated (Chang, 2005; Chang & Ho, 2009).

Chan and Ho (2009) indicated that WBLL provides more opportunities for individualized instruction with feedback than conventional classroom instruction. Son (2007) explored learner experiences in WBLL activities for English as a second language (ESL) learning. Analytical results confirmed that the Internet is a useful tool and supplementary resource for ESL learning. Moreover, participants who use Internet for ESL learning had positive attitudes toward WBLL, and indicated that they would like additional activities that could be completed in and outside class time (Son, 2008). Furthermore, some studies have demonstrated that WBLL can increase learner motivation and engage learners in culturally authentic and highly interactive language experiences (Chun & Plass, 2000; Mosquera, 2001). In short, compared with conventional CALL, WBLL provides learners with new and alternative ways of learning a language. Thus, this work designed a DRAS with SRL mechanisms for ESL learning and investigated its potential in promoting English-language e-reading performance.

**Digital reading annotation system**

Annotation can be used to summarize important ideas in an article and is an explicit expression of knowledge in the form of comments. Annotation also reveals the conceptual meanings of thoughts of annotators (Yang et al., 2011). People frequently have differing opinions about the same text; therefore, comments from different annotators may differ. Generally, a college freshman may be helped markedly by reading annotations from such experienced users as experts or senior students. Therefore, retaining and sharing comments is a key function of any effective annotation system. Frumkin (2005) indicated that if users can leave comments or annotations, this practice would open the door for sharing research experiences, facilitate collaborative research, and make it easy for future researchers to find materials they need in a particular collection.

Marshall (1997) argued that book annotations are useful to subsequent readers. Traditionally, annotating printed books by pencil or pen is the most common method of recording book-related knowledge, but has disadvantages when compared with knowledge storage in computers, knowledge dissemination, and knowledge sharing via the Internet. That is, using a digital annotation tool to annotate digital documents can overcome these shortcomings. Driss, Rachida and Amine (2006) designed a collaborative web annotation tool called SMARTNotes, which allows learners to conduct collaborative annotation for a digital reading article. Maristella and Nicola (2007) indicated that annotations can explain and enrich a resource with personal observations, and can be used to share ideas, thereby improving collaborative work practices. Yi et al. (2010) demonstrated that annotations are an effective means of interaction between students and teachers. Yang et al. (2007) developed a personalized annotation management system (PAMS) and evaluated how the PAMS can enhance knowledge sharing in online group reading activities.
They demonstrated that PAMS-enabled knowledge sharing improves the reading comprehension of students. Additionally, Mendenhall and Johnson (2011) applied an online annotation system to foster the development of critical thinking skills and reading comprehension of university undergraduates. To enhance the reading comprehension of ESL learners, this work develops a DRAS with SRL mechanisms that can facilitate individual learners to generate rich and high-quality annotations, thus helping them to improve the English-language reading comprehension via sharing their reading annotations.

Web-based self-regulated learning system

Self-regulated learning (SRL) was defined by Zimmerman as the degree to which learners are metacognitively, motivationally, and behaviorally active participants in their learning processes (Zimmerman, 1986a, 1986b). That is, SRL refers to a learning scenario in which learners set their own learning goals and plans, and then regulate and evaluate their own learning process (Narciss, Proske & Koerndle, 2007). Zimmerman, Bonner, and Kovach (1996) referred to various aspects of SRL in different studies when developing their SRL model, which contains the following four interrelated learning processes: self-evaluation and monitoring; goal setting and strategic planning; strategy implementation and monitoring; and, strategy outcome monitoring (Zimmerman, Bonner & Kovach, 1996). Notably, SRL helps learners self-examine and self-evaluate their learning performance by monitoring the learning goals they set during learning processes. Once learners set goals, they must be able to revise their learning strategies to achieve these goals.

Many recent studies have focused on the development of web-based learning systems with SRL mechanisms to promote web-based learning performance (Chen, 2009; Chang, 2005; Shih et al., 2010). Lee, Shen and Tsai (2008) experimentally determined whether web-based problem-based learning and SRL, or their combination, assist low-achieving students improve their computer skills in deploying application software. They demonstrated that the effects of SRL in promoting the computer skills of students were mostly positive and encouraging. Wang (2011) proposed a multiple-choice web-based assessment system, the Peer-Driven Assessment Module of the Web-based Assessment and Test Analysis (PDA-WATA) system, to help learners conduct SRL and to improve e-learning effectiveness. They demonstrated that the PDA-WATA system facilitates use of SRL behaviors and improved e-learning effectiveness. Shih et al. (2010) developed an SRL system with learning scaffolding supporting the independent learning skills of students. They indicated that SRL skills of students within a group with poor SRL abilities were significantly improved. Chang (2005) used a one-semester web-based course incorporating SRL strategies to help students improve their learning motivation. Analytical results demonstrated that a web-based environment with SRL strategies was helpful in keeping learning motivation so that students became increasingly confident in their understanding of course material and class performance. Our previous study (Chen, 2009) proposed a personalized e-learning system with SRL assistive mechanisms based on the SRL model proposed by Zimmerman, Bonner and Kovach (1996) to help learners enhance their SRL abilities for mathematical learning. Experimental results confirmed that the proposed SRL-assisted mechanisms helped learners accelerate their acquisition of SRL abilities in a personalized e-learning system, and promoted student learning performance. This work modifies the previous SRL assistive mechanisms (Chen, 2009) and embeds them into the proposed DRAS to help learners improve their reading comprehension and annotation abilities.

Research methodology

The proposed digital reading annotation system

Figure 1 shows the user interface of the proposed DRAS system providing several useful annotation functionalities that can annotate digital texts in the HTML format. These annotation functionalities include below

1. **Selection of annotation type:** When learners mark up a selected text, a popup menu with five options—word meaning, antonym, grammar, phrase and related links—appears to assist individual learners in annotating a digital text.
2. **Underlining:** This function can be used to identify and emphasize important points in a text.
3. **Browsing:** This function helps learners browse all annotations in an article.
4. **Voting:** Via this function, learners vote on the usefulness of annotations.
5. **Highlighting:** This function marks texts with annotations in **bright yellow**. When the mouse cursor is dragged
over highlighted text, the best annotations, which are identified based on voting, are shown in a popup window. When learners click on highlighted text, all annotations associated with the highlighted texts are displayed.

**Figure 1.** The user interface of the proposed DRAS system

**The SRL assistive mechanism in the proposed DRAS**

Although the proposed DRAS provides useful functionalities to help learners generate reading annotations, how to excite learners to contribute rich and high-quality reading annotations is obviously an essential issue. Consequently, functionalities of the DRAS are extended to include the SRL-assistive mechanisms that help learners self-observe and self-evaluate their learning processes, thus promoting learners’ spontaneous and autonomous abilities to increase the amount of reading annotations for enhancing ESL learning. This section describes the system components of the SRL mechanism embedded in the proposed DRAS as follows.

**Setting the self-monitor table**

While a learner uses the SRL mechanism for the first time, the self-monitor table menu is displayed (Fig. 2). In total, nine SRL indexes must be set in the self-monitor table for SRL. These SRL indexes include scheduled learning time (5–45 minutes), scheduled learning units (1–8 units), class ranking of the number of translation-type annotations, class ranking of antonym-type annotations, class ranking of phrase- and grammar-type annotations, class ranking of related link-type annotations (1–4), degree of effort (1–10), degree of concentration (1–10), and expected learning abilities (-3–3). After the self-monitor table is filled out, the proposed DRAS will guide learners entering the user interface for learning English-language texts based on contributing annotations and reading existed annotations. Moreover, when a learner logs into the system a second time, the system will show a record table of previous SRL outcomes as a reference.

**Self-regulated radar plot**

The radar plot of SRL has five SRL indicators, including four SRL competence indexes, and one learning performance index (Fig. 3). The four SRL competence indexes are the learning time index, effort for learning courseware index, reading rate index, concentrated learning index, and understanding index for learned courseware (Chen, 2009). In the self-regulated radar plot, blue points indicate a learner’s set values in the self-monitor table after the learner logs in the system. Conversely, red points represent a learner’s SRL status. The SRL radar plot can remind and motivate learners to achieve their learning goals.
Annotation ranking table

When a learner logs into the proposed DRAS, the upper left frame of the proposed system will show the annotation ranking table (Fig. 3), which aims to display the number of the four types of annotations including translation, antonym, grammar, and related link contributed by each learner and the ranking statues of the four types of annotations. A high level ranking status indicates good overall annotation performance. Based on the ranking level of the annotation type, the proposed system will show the corresponding encouraging words for learners for promoting them to contribute much more high-quality annotations.

Assessment scheme for reading annotation ability

Annotation types

According to differences in cognitive levels of annotation types, the proposed DRAS assigns different reward points to annotators who contribute different levels of annotations. Based on content analysis (Stemler, 2001), this work
divided annotations into two annotation levels—“basic annotation” and “advanced annotation.” To assess learner’s annotation ability, this work gave one reward point for “basic annotation” including word meaning and related links and two reward points for “advanced annotation” including antonym and grammar or phrase. Additionally, the validity of all annotations annotated by individual learners is confirmed by several senior English-language teachers. Invalid reading annotations do not receive reward points.

**Voting for high-quality annotations**

As is known, exploring high-quality annotations based on the collective intelligence of learner voting is helpful in recommending useful reading annotations to individual learners. High-quality reading annotations promote the reading comprehension of learners who read English-language texts online. To encourage learners to identify high-quality annotations, the proposed DRAS has a voting reward mechanism. In addition to encouraging learners to identify high-quality annotations, the reward mechanism also enhances the achievements of annotators, thereby encouraging learners to contribute high-quality annotations. Because each vote has the same value, 1 reward point is given for all annotation types.

**Evaluating the annotation ability of individual learners**

To investigate the relationship between annotation behaviors and reading comprehension, this work first calculates the total reward points based on the sum of the reward points of different levels of annotations and annotation voting, and then normalizes the total reward points based on the maximum total reward points to serve this value as the annotation abilities of individual learners. In other words, in addition to considering the frequency of annotations, the study also simultaneously considers annotation levels and whether annotations are valid to evaluate the annotation ability of individual learners.

**Reading materials, participants, and experimental procedures**

To provide articles for collaborative reading annotation, three reading units—each unit is composed of eight topics—were selected from the *English reading* textbook for Grade 7 students. All topics and posttest questions for assessing reading comprehension were designed by several senior English teachers at a junior high school in Taiwan. Additionally, this work applies the quasi-experiment nonequivalent control group design, as randomly selecting examinees as a research target is a difficult task in actual teaching scenarios. Experimental participants were randomly recruited from Grade 7 students in two classes at a junior high school in Taoyuan County, Taiwan. Each class has 32 students, 17 males and 15 females. The experimental group was randomly assigned from one of the two classes. The other class was assigned to the control group. The experimental group and control group used the proposed DRAS with and without the SRL mechanism for reading English-language texts, respectively. The learners of both the groups learnt the same English articles and performed self-directed learning without English teacher instruction during learning processes. The experimental procedures comprised an assessment of prior English proficiencies of both groups, performing cooperative reading annotations for the assigned English articles, learning the English articles with annotations, and an evaluation of reading comprehension and annotation abilities of both learner groups. The learning activities based on learning the English articles with annotations lasted 3 weeks. The annotated contents help readers obtain a deeper and broader understanding than when reading the English articles without annotations. Furthermore, compared to the learners of the control group, the learners of the experimental group will be immediately alerted by the SRL mechanisms during learning processes while their predetermined goals for SRL have not been achieved, thus cultivating their spontaneous and autonomous learning abilities as well as promoting their willingness to contribute annotations for cooperative ESL learning.

**Experimental results**

**Analysis of reading comprehension of both learner groups**

Before performing the reading annotation activity, three times English scores of school midterm exams are used to
confirm that both groups have the same English-language proficiencies based on the independent sample t-test. Next, this study determines whether reading comprehension of both groups differ significantly when using the DRAS with and without the SRL mechanism to support English-language texts reading. Table 1 shows analytical results, demonstrating that reading comprehension of both groups for the three reading units differed significantly (t = -2.584, p = .012 < .05; t = -2.473, p = .016 < .05; t = -2.702, p = .009 < .05). The experimental group had better reading comprehension than the control group, confirming that the reading comprehension of learners was significantly improved by the proposed DRAS with the SRL mechanism support.

Table 1. Statistical analysis of the reading comprehension performance of both groups

<table>
<thead>
<tr>
<th>Reading unit</th>
<th>Group</th>
<th>Number of learners</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CG</td>
<td>32</td>
<td>45.31</td>
<td>22.47</td>
<td>-2.584</td>
<td>.012*</td>
</tr>
<tr>
<td></td>
<td>EG</td>
<td>32</td>
<td>60.63</td>
<td>24.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unit 2</td>
<td>CG</td>
<td>32</td>
<td>45.00</td>
<td>18.18</td>
<td>-2.473</td>
<td>.016*</td>
</tr>
<tr>
<td></td>
<td>EG</td>
<td>32</td>
<td>56.72</td>
<td>19.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unit 3</td>
<td>CG</td>
<td>32</td>
<td>47.34</td>
<td>21.48</td>
<td>-2.702</td>
<td>.009*</td>
</tr>
<tr>
<td></td>
<td>EG</td>
<td>32</td>
<td>62.19</td>
<td>22.47</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* indicates p < .05

Next, the effect of gender on reading comprehension between both groups is assessed. Tables 2 and 3 show analytical results based on the independent sample t-test. The results show that reading comprehension of males in both groups did not differ significantly (t = -1.188, p = .243 > .05; t = -1.823, p = .078 > .05; t = -1.286, p = .208 > .05). For the female learners, except for reading comprehension for the second reading unit, reading comprehension for the first and third units differed significantly (t = -2.708, p = .011 < .05; t = -1.648, p = .111 > .05; t = -2.636, p = .014 < .05). This indicates that the females in the experimental group using the proposed DRAS with the SRL mechanism had better reading comprehension than the females in the control group. That is, a gender difference in reading comprehension existed when using the DRAS with and without the SRL mechanism support. Obviously, the proposed DRAS with the SRL mechanism support promotes reading comprehension of female learners more than that of male learners.

Table 2. Statistical analysis of reading comprehension performance of the male learners of both groups

<table>
<thead>
<tr>
<th>Reading unit</th>
<th>Group</th>
<th>Number of males</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit 1</td>
<td>CG</td>
<td>17</td>
<td>43.53</td>
<td>22.55</td>
<td>-1.188</td>
<td>.243</td>
</tr>
<tr>
<td></td>
<td>EG</td>
<td>17</td>
<td>53.82</td>
<td>27.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unit 2</td>
<td>CG</td>
<td>17</td>
<td>42.35</td>
<td>15.72</td>
<td>-1.823</td>
<td>.078</td>
</tr>
<tr>
<td></td>
<td>EG</td>
<td>17</td>
<td>54.41</td>
<td>22.28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unit 3</td>
<td>CG</td>
<td>17</td>
<td>48.82</td>
<td>21.62</td>
<td>-1.286</td>
<td>.208</td>
</tr>
<tr>
<td></td>
<td>EG</td>
<td>17</td>
<td>59.12</td>
<td>24.95</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Statistical analysis of reading comprehension for the female learners of both groups

<table>
<thead>
<tr>
<th>Reading unit</th>
<th>Group</th>
<th>Number of females</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit 1</td>
<td>CG</td>
<td>15</td>
<td>65.800</td>
<td>24.16</td>
<td>-2.708</td>
<td>.011*</td>
</tr>
<tr>
<td></td>
<td>EG</td>
<td>15</td>
<td>75.507</td>
<td>16.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unit 2</td>
<td>CG</td>
<td>15</td>
<td>47.33</td>
<td>22.98</td>
<td>-1.648</td>
<td>.111</td>
</tr>
<tr>
<td></td>
<td>EG</td>
<td>15</td>
<td>68.33</td>
<td>19.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unit 3</td>
<td>CG</td>
<td>15</td>
<td>48.00</td>
<td>20.77</td>
<td>-2.636</td>
<td>.014*</td>
</tr>
<tr>
<td></td>
<td>EG</td>
<td>15</td>
<td>59.33</td>
<td>16.68</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* indicates p < .05
Analysis of reading annotation abilities of both learner groups

This section assesses whether annotation abilities of both groups differ significantly when using the DRAS with and without the SRL mechanism to support English-language texts reading. Table 4 shows analytical results, demonstrating that the reading annotation ability of both learner groups differed significantly for all reading units ($t = -4.402, p = .000 < .05; t = -3.038, p = .004 < .05; t = -2.640, p = .010 < .05$), and the annotation ability of the experimental group was superior to that of the control group. That is, the reading annotation ability of experimental group learners was significantly enhanced when using the DRAS with the SRL mechanism support.

Moreover, the effects of the gender difference in reading annotation ability between the control group and experimental group are also assessed. Tables 5 and 6 show independent sample $t$-test results for males and females in both groups, respectively. The results show that the reading annotation ability of male learners in both groups for the three units differed significantly ($t = -3.737, p = .001 < .05; t = -3.009, p = .007 < .05; t = -3.170, p = .003 < .05$), and male learners in the experimental group was superior to that of male learners in the control group. Conversely, the reading annotation ability of female learners of the experimental group for the first unit was significantly different from that of female learners of the control group ($t = -2.425, p = .022 < .05$), and that of experimental group was superior to that of the control group. However, the reading annotation ability of female learners in the second and third units did not differ significantly ($t = -1.027, p = .313 > .05; t = -3.979, p = .694 > .05$). Therefore, a gender difference in reading annotation ability existed when using the proposed DRAS with and without the SRL mechanism. Obviously, the proposed DRAS with the SRL mechanism was more helpful in promoting the reading annotation ability of male learners than those of female learners.

### Table 4. Statistical analysis of the reading annotation abilities of both groups

<table>
<thead>
<tr>
<th>Reading unit</th>
<th>Group</th>
<th>Number of learners</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>$t$</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit 1</td>
<td>CG</td>
<td>32</td>
<td>14.34</td>
<td>13.25</td>
<td>-4.402</td>
<td>.000***</td>
</tr>
<tr>
<td></td>
<td>EG</td>
<td>32</td>
<td>34.53</td>
<td>22.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unit 2</td>
<td>CG</td>
<td>32</td>
<td>14.16</td>
<td>11.15</td>
<td>-3.038</td>
<td>.004**</td>
</tr>
<tr>
<td></td>
<td>EG</td>
<td>32</td>
<td>27.28</td>
<td>21.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unit 3</td>
<td>CG</td>
<td>32</td>
<td>16.41</td>
<td>12.30</td>
<td>-2.640</td>
<td>.010*</td>
</tr>
<tr>
<td></td>
<td>EG</td>
<td>32</td>
<td>27.94</td>
<td>21.43</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* indicates $p < .05$; ** indicates $p < .01$; *** indicates $p < .001$

### Table 5. Statistical analysis of the reading annotation abilities of the male learners of both groups

<table>
<thead>
<tr>
<th>Reading unit</th>
<th>Group</th>
<th>Number of males</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>$t$</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit 1</td>
<td>CG</td>
<td>17</td>
<td>14.76</td>
<td>13.76</td>
<td>-3.737</td>
<td>.001**</td>
</tr>
<tr>
<td></td>
<td>EG</td>
<td>17</td>
<td>38.76</td>
<td>22.62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unit 2</td>
<td>CG</td>
<td>17</td>
<td>12.82</td>
<td>10.58</td>
<td>-3.009</td>
<td>.007**</td>
</tr>
<tr>
<td></td>
<td>EG</td>
<td>17</td>
<td>33.35</td>
<td>26.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unit 3</td>
<td>CG</td>
<td>17</td>
<td>13.12</td>
<td>8.59</td>
<td>-3.170</td>
<td>.003**</td>
</tr>
<tr>
<td></td>
<td>EG</td>
<td>17</td>
<td>32.76</td>
<td>24.07</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** indicates $p < .01$

### Table 6. Statistical analysis of the reading annotation abilities of the female learners of both groups

<table>
<thead>
<tr>
<th>Reading unit</th>
<th>Group</th>
<th>Number of females</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>$t$</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit 1</td>
<td>CG</td>
<td>15</td>
<td>13.87</td>
<td>13.11</td>
<td>-2.425</td>
<td>.022*</td>
</tr>
<tr>
<td></td>
<td>EG</td>
<td>15</td>
<td>29.73</td>
<td>21.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unit 2</td>
<td>CG</td>
<td>15</td>
<td>15.67</td>
<td>11.94</td>
<td>-1.027</td>
<td>.313</td>
</tr>
<tr>
<td></td>
<td>EG</td>
<td>15</td>
<td>20.40</td>
<td>13.27</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Analysis of the difference in reading comprehension and reading annotation ability of the experimental group learners with different SRL abilities

To assess how different SRL abilities affect reading comprehension and annotation ability in the experimental group, the SRL indexes, including the achievement index of learning time, achievement index of effort level in learning courseware, achievement index of reading rate, achievement index of concentrated learning, and degree of understanding of learned courseware proposed in our previous work (Chen, 2009), are used to categorize learners in the experimental group into a good or poor SRL group. Table 7 shows comparison results based on the independent sample t-test. Analytical results show that the reading comprehension of the good SRL group and poor SRL group differed significantly ($t = 6.267, p = .000 < .05; t = 6.517, p = .000 < .05; t = 5.695, p = .000 < .05; t = 7.163, p = .000 < .05$) for the three units, and the good SRL group was superior to that of the poor SRL group.

Table 7. Comparison of the reading comprehension performances of the experimental group learners with different SRL abilities

<table>
<thead>
<tr>
<th>Reading unit</th>
<th>Group</th>
<th>Number of learners</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit 1</td>
<td>Good SRL</td>
<td>16</td>
<td>79.06</td>
<td>17.48</td>
<td>6.267</td>
<td>.000***</td>
</tr>
<tr>
<td></td>
<td>Poor SRL</td>
<td>16</td>
<td>42.19</td>
<td>15.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unit 2</td>
<td>Good SRL</td>
<td>16</td>
<td>71.56</td>
<td>14.69</td>
<td>6.517</td>
<td>.000***</td>
</tr>
<tr>
<td></td>
<td>Poor SRL</td>
<td>16</td>
<td>41.88</td>
<td>10.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unit 3</td>
<td>Good SRL</td>
<td>16</td>
<td>78.13</td>
<td>14.93</td>
<td>5.695</td>
<td>.000***</td>
</tr>
<tr>
<td></td>
<td>Poor SRL</td>
<td>16</td>
<td>46.25</td>
<td>16.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>Good SRL</td>
<td>16</td>
<td>76.24</td>
<td>13.98</td>
<td>7.163</td>
<td>.000***</td>
</tr>
<tr>
<td></td>
<td>Poor SRL</td>
<td>16</td>
<td>43.44</td>
<td>11.83</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
*** indicates $p < .001$

Additionally, how different SRL abilities affect the reading annotation ability of learners in the experimental group is also investigated. Table 8 shows analytical results, indicating that the reading annotation ability of experimental group learners for the first and second units did not differ significantly ($t = 1.392, p = .174 > .05; t = 1.699, p = .104 > .05$); however, the reading annotation ability of experimental group learners in the third unit and means of the three units differed significantly ($t = 2.391, p = .023 < .05; t = 2.083, p = .046 < .05$). These comparison results partially demonstrate that reading annotation ability of the good SRL group was superior to that of poor SRL group.

Table 8. Comparison of the reading annotation abilities of the experimental group learners with different SRL abilities

<table>
<thead>
<tr>
<th>Reading unit</th>
<th>Group</th>
<th>Number of learners</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit 1</td>
<td>Good SRL</td>
<td>16</td>
<td>39.94</td>
<td>22.27</td>
<td>1.392</td>
<td>.174</td>
</tr>
<tr>
<td></td>
<td>Poor SRL</td>
<td>16</td>
<td>29.13</td>
<td>21.68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unit 2</td>
<td>Good SRL</td>
<td>16</td>
<td>33.63</td>
<td>27.15</td>
<td>1.699</td>
<td>.104</td>
</tr>
<tr>
<td></td>
<td>Poor SRL</td>
<td>16</td>
<td>20.94</td>
<td>12.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unit 3</td>
<td>Good SRL</td>
<td>16</td>
<td>36.38</td>
<td>25.00</td>
<td>2.391</td>
<td>.023*</td>
</tr>
<tr>
<td></td>
<td>Poor SRL</td>
<td>16</td>
<td>19.50</td>
<td>13.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>Good SRL</td>
<td>16</td>
<td>36.65</td>
<td>22.45</td>
<td>2.083</td>
<td>.046*</td>
</tr>
<tr>
<td></td>
<td>Poor SRL</td>
<td>16</td>
<td>23.19</td>
<td>12.80</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
* indicates $p < .05$
Correlation analysis between reading comprehension and reading annotation ability in both learner groups

Pearson correlation analysis is applied to assess the strength of correlations between reading comprehension and reading annotation ability for both groups. The correlation coefficient between reading comprehension and reading annotation ability for the experimental group learners was .235 ($p = .021 < .05$). However, the correlation between reading comprehension and reading annotation ability in the control group was not statistically significant. Obviously, using the DRAS with the SRL mechanism to support reading English-language texts online is the key factor that builds the correlation between the reading comprehension and reading annotation abilities of the experimental group learners although the correlation coefficient is only weakly positive.

Discussion

First of all, this work confirmed the proposed DRAS with the SRL mechanism efficiently promoted reading comprehension of learners who set learning goals and self-monitored their progress for reading English-language texts online. Many studies have indicated that SRL strategies can help learners improve their learning performance (Boekaerts, 1997; Pintrich, 2000; Zimmerman & Schunk 2001; Chang, 2005; Chen, 2009; Shih et al., 2010; Cheng, 2011). Particularly, Cheng (2011) argued that the SRL abilities of students, including learning motivation, goal setting, action control, and learning strategies, have significant effects on learning performance. Moreover, Lei et al. (2001) indicated that setting learning goals can help students understand learning tasks during SRL processes. Study results of this work are consistent those obtained by the above studies. Additionally, experimental results show that the proposed DRAS with the SRL mechanism efficiently enhanced the reading annotation ability of learners by setting reading annotation goals and displaying the rank of reading annotations. Observations indicate that learners frequently cared about the rank of their annotations via the annotation reward mechanism. Therefore, most learners were encouraged to annotate articles, such that high-quality annotations were generated, thereby promoting the reading comprehension of other learners. This work also found that most learners had a positive attitude toward this collaborative learning behavior. This analytical result is similar to that obtained by the study by Chen and Chang (2012), which explored social ranking of individual learners in a cooperative problem-based learning (PBL) environment that can encourage learners to interact actively with learning peers.

Moreover, gender differences in education have been recognized as an important research focus for a long time (Lee, 2002; Yukelturk & Bulut, 2009). Several previous studies claimed that male and female learners use SRL strategies during their learning processes differently (Lee, 2002; Wolters & Pintrich, 1998; Zimermann & Martinez-Pons, 1990). Zimermann and Martinez-Pons (1990) determined that girls tend to employ self-monitoring, goal setting, planning, and structuring of their study environment more often than boys. Wolters and Pintrich (1998) confirmed that gender differences in motivational and cognitive strategies existed during SRL. Lee (2002) also indicated that gender differences in motivational and behavioral learning strategy components are evident in cyber-learning contexts. Analytical results obtained by this work also confirmed that gender differences in reading comprehension and reading annotation ability existed when using the DRAS with and without the SRL mechanism for supporting English-language texts reading. The proposed DRAS with the SRL mechanism was more helpful in promoting the reading comprehension of female learners than that of male learners; however, the proposed DRAS with the SRL mechanism was more helpful in promoting reading annotation abilities of male learners than that of female learners. In addition to that gender difference may affect SRL performance, there are the other factors outside the research scope of the study to affect SRL performance. Pintrich, Roeser and De Groot’s study (1994) confirmed that motivational beliefs (self-efficacy, intrinsic value, test anxiety) was positively related to higher levels of SRL. Moreover, Schapiro and Livingstone’s study (2000) suggested that the natural dynamic component, reflecting qualities such as curiosity, enthusiasm, willingness to take risks, and persistence, actually underlies and drives the strategic behavior of SRL.

Moreover, experimental results show that significant differences existed in reading comprehension and reading annotation abilities of learners with good and poor SRL abilities in the experimental group. Reading comprehension and reading annotation abilities of experimental group learners with good SRL ability were higher than those of learners with poor SRL ability. Young (1996), who examined the effect of SRL strategies on learning performance in learner-controlled and program-controlled computer-based instruction (CBI), found that performance differences between learners with good SRL strategies and those with poor SRL strategies were greatest under learner control. This work found that the reading annotation ability of learners in the experimental group was positively correlated
with reading comprehension; however, no such correlation existed in the control group. These analytical results also prove that the proposed DRAS with the SRL mechanism encouraged learners to contribute high-quality annotations, thereby enhancing the reading comprehension of other learners. Furthermore, the SRL abilities of learners in the experimental group were positively correlated with their reading comprehension, indicating that reading comprehension of learners in the experimental group can be evaluated according to their SRL indexes assessed by the proposed system.

Finally, although the proposed DRAS with SRL mechanism provides benefits on ESL learning, some limitations of this study merit further consideration. First, the SRL indexes which include the scheduled learning time, degree of effort, and degree of concentration are evaluated in accordance with valid learning time identified by detecting the operations mouse and keyboard actions during a specific amount of time. This leads to that the valid learning time is not completely accurate. Second, although the proposed DRAS with SRL mechanism support positively affects learners to contribute rich reading annotations, the quality of learner generated annotations cannot be guaranteed. This may affect learner’s reading performance.

**Conclusions and future works**

This study investigates the effects of the proposed DRAS with the SRL mechanism, which was applied to support online English-language texts reading, on reading comprehension performance and reading annotation ability for Grade 7 students at a junior high school in Taiwan. Based on statistical analyses, several major findings are summarized as follows. First, compared with control group learners, the reading comprehension and reading annotation ability of experimental group learners who used the proposed DRAS with the SRL mechanism for English-language reading improved significantly. Moreover, gender differences in reading comprehension and reading annotation ability existed when using the proposed DRAS with and without the SRL mechanism for English-language reading. The proposed DRAS with the SRL mechanism promoted the reading comprehension of female learners more than that of male learners; however, the proposed DRAS increased the reading annotation ability more for male learners than for female learners. Additionally, reading comprehension and reading annotation abilities of experimental group learners with high SRL ability were all higher than those of learners with poor SRL ability. Further, the relationship between reading comprehension and reading annotation ability of experimental group learners was weakly positive, and no statistically significant correlation existed in the control group.

Finally, several suggestions for future works are addressed based on experimental results and participant responses. First, to determine SRL indexes accurately, the proposed DRAS can consider assessing valid learning time based on learning attention potentially identified by the brainwave system (Haapalaainen, Kim, Forlizzi, & Dey, 2010). Second, to guarantee the quality of learner contributed annotations, developing an intelligent mechanism which can automatically filter out poor-quality annotations based on collective intelligence or data mining should be considered.

**References**


Mathematics Synchronous Peer Tutoring System for Students with Learning Disabilities

Mengping Tsuei
Graduate School of Curriculum and Instructional Communications Technology, National Taipei University of Education, Taipei, Taiwan // mptsuei@gmail.com

(Submitted August 24, 2012; Revised March 11, 2013; Accepted May 10, 2013)

ABSTRACT

The purpose of this study was to develop and explore the impact of a synchronous peer tutoring system, which integrated a structured peer tutoring strategy with technological advances, for students with learning disabilities (LD). The system provided a learning activity management module for teachers to administer peer tutoring activities, and included various math representation objects, communication tools, and reward schema for facilitating online peer learning activities. Four fourth-grade students with LD participated in this study in the online peer learning setting for two semesters. The results indicated that the proposed system was effective to enhance the mathematics learning of students with LD, especially the learning of conceptual and application math problems. Students with LD showed improvement in math fluency on conceptual problems. The findings also supported the effectiveness of the math objects provided by the synchronous peer tutoring system for facilitating communication among students and understanding of math concepts in the online activities. The results are discussed from the perspective of online peer learning research.

Keywords

Synchronous system, Peer tutoring, Elementary school, Math learning, Learning disabilities

Introduction

Peer tutoring is one of the most well-studied strategies in mathematics instruction. The structured guidance format of peer tutoring effectively promotes helping behaviour in peer-mediated math learning, especially for students who have difficulty with the material (Webb & Mastergeorge, 2003). Peer tutoring has been defined as “people from similar social groupings who are not professional teachers, helping each other to learn, and learning themselves by teaching” (Topping & Ehly, 1998). Peer tutoring has often been employed to enhance academic skills within the inclusive classroom setting (Mastropoeri et al., 2001). Greenwood, Delquadri, and Carta (1997) demonstrated that elementary students with special needs showed improved performance in basic reading, spelling, and mathematics through class-wide peer tutoring activities in general educational classes. Fuchs, Fuchs, Yazdian, and Powell (2002) studied peer-assisted learning strategies in the inclusive classroom for elementary children. They found that students with disabilities showed significantly greater progress than without disabilities.

In mathematics learning, most of the research conducted has supported the conclusion that tutoring results in a greater improvement in lower-order skills compared to mixed outcomes. Calhoon and Fuchs (2003) found that peer-assisted learning strategies improved computational math skills for secondary school students with disabilities. However, no significant difference was found in terms of concept or application math skills or on the state-wide test. Schloss, Kobza, and Alper (1997) indicated that peer tutoring improved currency skills for students with moderate mental retardation. Although these are important findings, these studies only targeted basic skills. As Bryant, Bryant, and Hammill stated (2000), multistep and word problems and the skills associated with them are the most problematic topics for students with disabilities. Thus, the effectiveness of peer tutoring on mathematics learning for students with learning disabilities (LD) still needs to be investigated.

There are some pedagogical advantages of peer tutoring for students with LD, who have been characterized as passive learners in the classroom (Hallahan, Kauffman, & Lloyd, 1996). First, the peer-assisted learning environment ensures active participation, which is typically lacking in instruction for students with LD (Limbrk, McNagghton, & Glynn, 1985). Second, the effectiveness of this intervention allows students to receive individual attention and immediate feedback (Stenhoff & Lignugaris/Kraft, 2007). However, the greatest challenge in peer tutoring procedures for students with LD is that they may have problems with expressive communication skills. Students with LD exhibit limited learning behaviours, such as asking questions (Wood & Algozzine, 1994). They often experience
difficulty determining what to say, remembering how to say it, and saying it aloud in front of others (Schott & Windsor, 2000). Therefore, the scaffolding tools for facilitating online tutoring for LD students are needed.

Mathematical learning in primary education is mostly through the observation of interactions with external information (e.g., real objects, physical manipulatives, and representations). Researchers have noted the importance of adapting the instructional material to the needs of students with disabilities, such as with picture-based icons or symbolic representatives (Harrison, 2002). The use of visual and concrete representations facilitated student understanding of mathematical problem solving (Jitendra et al., 1998). Maccini and Gagnon (2000) identified manipulatives as exceptionally useful for students with high-incidence disabilities in mathematics. Use of these concrete aids has been determined to be an effective medium for students across grade and developmental levels, including students with disabilities (Cass, Cates, Smith, & Jackson, 2003).

Computer-based learning environments provide multiple representations of transformations, supporting cognitive symbol translations through various representations (Hwang, Huang, & Dong, 2009). Moyer (2001) stated that the use of virtual manipulatives with a graphical user interface provided interactive and visual representations of dynamic objects for constructing mathematical knowledge. Hwang et al. (2009) developed the multimedia whiteboard system for children to learn mathematics. The system had the drawing tools and editing functions. The solution and criticizing content can be stored automatically on the website. Students and teachers can discuss the math solutions in face-to-face way. However, the system only supported students to solve math problems individually and provided limited mathematics objects. Kolloffel, Eysink, and Jong (2011) found that by providing representational tools in collaborative learning settings, the learning results were significantly higher than those in individual settings.

The CSCL environment was found to be effective for introducing peer tutoring on an organized educational basis (Smet, Keer, Wever, & Valcke, 2010; Yip, 2004). Yip (2004) developed a web-based system to support peer tutoring activities for university students, which facilitated interaction among students and expanded face-to-face tutoring discussions. Smet, Keer, Wever, and Valcke (2010) explored the influence of three tutor training types on the characteristics of fourth year students in asynchronous discussions. The results indicated that it may be fruitful to provide guidelines to novice tutors through a specific training approach. With new technological advances in network computing and theories of collaborative learning, researchers now support the use of a synchronous approach. Kong (2008) developed a cognitive tool for supporting the teaching and learning of fractions in primary mathematics classrooms. The results found that students learned fraction concepts by graphical presentations, representations, or operations in computer-supported learning environments, which helped them to develop conceptual understanding and procedural knowledge (Kong, 2008). However, a major drawback of these systems is that the math activities are only limited to specific content areas and cannot be incorporated to the existing curriculum. Xenos, Avouris, Stavrinoudis, and Margaritis (2009) showed that synchronous peer tutoring was positively received by their participants. However, research about online synchronous peer tutoring, especially for elementary children (Tsuei, 2011), remains rather scarce (Cheng & Ku, 2009).

Given the lack of studies for elementary students with LD in the online settings, the aim of the current study was to develop a synchronous CSCL system that provided structured peer tutoring strategies and virtual manipulatives to promote mathematics learning for students with LD. On the basis of the above research, the present study investigated the primary question: What are the effects of the synchronous peer tutoring strategy on mathematical learning for students with LD? This paper presents the development of the synchronous peer tutoring system and the results of the experimental study.

The synchronous peer tutoring system

The synchronous peer tutoring system proposed in this study, called the G-Math Peer Tutoring System (G-Math), was developed by using Massive Multiplayer Online Game (MMOG) technology that provided multiuser dungeons (MUDs). The MMOG technology adds a new dimension to game play, allowing for social interaction between humans (Gran & Reinemo, 2008). The multiuser server GATE was used to develop G-Math for facilitating interactions among connected users. G-Math was also developed with web technologies, including the Adobe Flash interface and PHP, MySQL, and XML.
The framework of G-Math included two subsystems (Figure 1). The mathematics learning activities management system was developed for teachers to administer peer tutoring activities. The peer tutoring system allowed children to perform the “game-like” computer-mediated peer learning in face-to-face classrooms.

Figure 1. The framework of the synchronous peer tutoring system

The learning activity management system was a tool for teachers to administer the peer tutoring activities. The grouping module was used to assign specific students into the same group and to assign the initial role of tutor and tutee. The mathematics tutoring activities module was used to assign different mathematics problem sets from the item bank to specific groups. Teachers could assign different learning units (e.g., fractions, whole-integer computation, and geometry), math problem types, and difficulty levels to different dyad. According to these parameters, the system randomly selected mathematics questions from item bank for students to implement the peer tutoring online.

Figure 2 shows the interface of the peer tutor system. Information about peers and personal information were shown on the right side of the screen. The G-Math system provided the personal information, including the name, experience-value, avatar, and awarded-scores, of the student (Figure 2-1). Awarded-scores were the aggregated points obtained from the tutor and other members of the same group, determined by their performance of solving math questions during the peer tutoring activities (Figure 2-2). To enhance student motivation, the experience-value was increased by answering more math problems correctly. If students answer a math problem correctly, they get one point of experience-value. The system automatically calculates ten points of experience-value as one experience level. Experience levels were shown as symbols on the sides of the avatar for each student (Figure 2-3). Symbols were represented by different coloured pictures of hats, roller skates, and kick scooters. Information about peers included the icons and names of avatars in the same group. When the student logged in, his or her icon of avatar became a bright colour (Figure 2-4).

The middle screen was the peer tutoring whiteboard, which was the synchronous peer tutoring section for the all members in the same group. On top of this section, the whiteboard showed the mathematics problem, which was generated automatically by the system. These math problems are assigned by the teacher using the learning activity management system (Figure 2-5). As the student solved the mathematics, he or she could use the virtual manipulatives and mathematics symbols on the top left of the screen to express his or her ideas, annotations, and mathematical reasoning (Figure 2-6). Based on the fourth-grade math curriculum, we designed the mathematics learning objects for students to solve math problems. G-Math provided 13 categories of 220 math learning objects for children to manipulate, i.e., computation symbols, integer symbols, decimal symbols, fraction symbols, measuring tools (Figure 2-7, Figure 3). Students could manipulate these symbols themselves by dragging, drawing, scaling, and rotating them. For example, the student tutors can explain how they estimate angle measurements by dragging the triangle object. All of students in the same group could view other members manipulating these objects in the tutoring whiteboard area.

Communication scaffolding tools were located at the bottom of the screen. G-Math provided various communication scaffolding tools to facilitate the tutoring process in math peer learning. The communication scaffolding tools in the face-to-face settings can facilitate the tutor’s and tutee’s interactive behaviours. The tutor could use guided sentences in the chatting area for peer instruction, task coordination, and feedback (Figure 2-8). Guided sentences were the frequently used ones by tutors (Figure 2-9), e.g., “This is a very important step for solving this problem.” and “Do you need some help?” and “Good job.” The sentences for asking questions were also provided for the tutee, e.g., “What is the meaning of this equation?” and “I don’t get it.” The feature designed in the face-to-face online peer tutoring scenario is to facilitate the help-seeking and help-giving behaviours of elementary students, especially with
learning disabilities. This feature also reduced typing barriers for elementary students. Students can also use emoticons to give feedbacks (Figure 2-10). Emoticons, which were displayed on the right side of the icon of the avatar, were provided to enhance the online interactions of students.

Figure 2. Interface of the synchronous peer tutoring system: (1) name and avatar of the user, (2) scores, (3) experience value and levels, (4) group members, (5) mathematics problem, (6) mathematics objects and tutoring area, (7) math learning objects, (8) chatting area, (9) guided sentences, and (10) emoticons

Figure 3. Examples of math learning objects: (A) fraction symbols (B) measuring tools (C) general symbols (D) integer symbols

Methods

Participants

This study was implemented in the resource classroom in an elementary school in Taipei. Four fourth-grade students who received mathematics remedial instruction participated in this study. All of subjects were identified by the Committee Responsible for Identification and Placement of Gifted and Disabled Students in Taipei city. According to the committee, the discrepancy model is one of the commonly adopted criteria to evaluate whether a student is eligible for special education services. The LD students’ severe discrepancies between intellectual ability and academic achievement exist in one or more academic areas. The subject A, B and C were identified as LD students. Subject D was also identified as having mild mental retardation.

The face-to-face online peer tutoring on G-Math was implemented at a rate of one session (40 minutes) per week for two semesters (one year). Students were paired in the G-Math peer tutoring activities on the basis of the remedial mathematics lessons that the teacher assigned. The teacher reassigned the groups biweekly.
Face-to-face online peer tutoring activities

The face-to-face online peer tutoring section was implemented once a week. First, the teacher taught the concepts of mathematics problems for 10 minutes. Then, four students in the resource classroom were seated in front of computers to use G-Math for 30 minutes. Two students in the same group were seated in front of the computer side by side for peer tutoring. In this study, the turn-taking strategy of peer tutoring was adopted. The procedures of peer tutoring session were as following:

Step 1. Before the class, the teacher assigned the math unit for the peer learning activities.
Step 2. The tutor notified the group member to start the tutoring session by pressing the “start” button.
Step 3. Then, the tutor pressed the “next question” button, a math question was shown on the whiteboard in the middle of the screen for the tutor to solve. He or she served as a model of problem-solving for the tutee.
Step 4. The tutee watched or asked questions about the solving process when the tutor solved the problem. He or she gave scores to the tutor according to the performance.
Step 5. To promote metacognitive thinking processes, G-Math provided the correct answers for students to reflect his or her solutions. Students could gain one point of experience-value by solving a problem correctly.
Step 6. The tutor assigned the next math problem for the tutee by clicking the tutee’s avatar icon. Now, their roles were exchanged. The tutee became the tutor. The tutoring procedures turned back to the step 3 till the end of the class.

In the above procedures, students used the mathematics objects and symbols in the manipulatives area to answer the math problem in the tutoring area. For example, students used the money objects to explain how to solve the integer division problem. They could use the place value model to help them solve the decimal subtraction problems.

In addition to discussing concepts face-to-face, they could use the communication scaffolding tools to pose questions or give feedback. For example, the tutee can use the guided sentence “Please explain again!” for asking questions.

The teacher observed the students’ tutoring behaviours and corrected misconceptions as needed.

Instrument

The mathematics curriculum-based measurement (CBM) instrument was used to measure and monitor student mathematics proficiency. The CBM is a database problem-solving model for indexing the academic competence and tracking progress of students through ongoing assessment (Deno, 1985). A web-based CBM system (ECBM) was used in this study to measure and monitor the mathematics proficiency of students. ECBM included all question types in the mathematics textbooks at every grade level. The ECBM randomly selected 10 math questions from the item bank for the specific grade level as a math CBM probe. There were 10 math questions in a mixed-problem math probe, including five conceptual, three computational, and two application questions. The CBM probes supported the adequacy of the reliability ($r = .63-.76, p < .05$) and validity ($r = .40-.84, p < .05$) (Tsuei, 2008). Three CBM probes were administered for the baseline scores in each semester. A CBM probe was administered once a week before each peer tutoring section. There were 19 CBM probes implemented each semester in this study. Digital scoring was implemented to evaluate student performance on the CBM probe. Scores were calculated by adding the number of correct digits (Tsuei, 2008).

Results

Changes in student scores on the mathematics CBM over one year

Mean differences of the total scores for each student and for different problem types across the two semesters were compared with non-parametric tests. The Mann-Whitney U-tests were carried out. The Rosenthal’s $r$ was used to calculate the effect size (ES) (Rosenthal, 1991). The results for every subject were shown in Table 1. Subjects A, B, and C showed significant improvements in mathematics in the G-Math peer tutoring learning environment (ES $r = -0.37, -0.53, -0.33$). This represented a moderate effect for subject A and C (above 0.3 criterion for a medium effect size) and large effect (the effect size above .5) for subject B. Subject D also showed a slight improvement ($r = -0.13$).
Table 1. CBM scores of students by semester

<table>
<thead>
<tr>
<th>Subject</th>
<th>Mean1 (SD)</th>
<th>Mean2 (SD)</th>
<th>Mann-Whitney U</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>19.89 (8.85)</td>
<td>28.32 (11.44)</td>
<td>103.50</td>
<td>-2.25*</td>
</tr>
<tr>
<td>B</td>
<td>12.00 (13.54)</td>
<td>27.32 (14.07)</td>
<td>68.50</td>
<td>-3.28**</td>
</tr>
<tr>
<td>C</td>
<td>24.00 (9.53)</td>
<td>35.32 (18.88)</td>
<td>120.50</td>
<td>-1.78*</td>
</tr>
<tr>
<td>D</td>
<td>14.00 (7.77)</td>
<td>15.37 (7.97)</td>
<td>154.00</td>
<td>-0.78</td>
</tr>
</tbody>
</table>

*p < 0.05, **p < 0.01

The results of the Mann-Whitney U-tests across problem types (conceptual, computational, and application problems) on the CBM are shown in Table 2. Subjects A, B, and C showed a significant growth in terms of the conceptual problem type ($r = -0.37, -0.40, -0.68$). These findings indicated the medium and large effect size. No subject showed significant improvement on the computational problem-type. Only subject B showed a significant improvement on the application problem-type ($r = -0.50$). It was worthy to note that the subject D’s application problem-type data showed a moderate effect size ($r = -0.20$).

Table 2. CBM scores by semester and problem type

<table>
<thead>
<tr>
<th>Problem types</th>
<th>Conceptual questions</th>
<th>Computational questions</th>
<th>Application questions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean1 (SD)</td>
<td>Mean2 (SD)</td>
<td>Mann-Whitney U</td>
</tr>
<tr>
<td>A</td>
<td>3.84 (4.96)</td>
<td>2.68 (2.61)</td>
<td>-2.31*</td>
</tr>
<tr>
<td>B</td>
<td>9.63 (9.92)</td>
<td>7.89 (6.89)</td>
<td>95.50</td>
</tr>
<tr>
<td>C</td>
<td>102.00</td>
<td>95.50</td>
<td>6.88 (9.36)</td>
</tr>
<tr>
<td>D</td>
<td>3.45 (5.77)</td>
<td>3.80 (8.46)</td>
<td>-4.18***</td>
</tr>
</tbody>
</table>

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Growth rate of students on mathematics CBM

The growth rate (slope) played an important role in CBM determination (Deno, 1985). The instructor may use the slope as the one in the regression equation to expect a student’s CBM score on a specific day or week. By comparing individual student’s slope (trend line) and the goal line, teachers obtained valuable information about instructional adjustments. As comparing the growth rates of the students with and without learning disabilities, teachers can determine whether students with special services (i.e., resource classroom) can be back to general classroom (Shinn 1998).

Results of the mathematics CBM graphs of the students for the first and second semesters were generated with the ECBM system (Figures 4-7). Each graph used three colours to depict CBM performance. The baseline (in red) was drawn from the scores of three pre-tests. The median score was chosen as the initial performance level for the child. The goal line (in green), which the child was expected to achieve during the monitoring period, was drawn from the initial score of the child connected with his or her weekly growth scores in math. The expected growth rate was based on the normative growth rate achieved by students at the same grade level. The blue line was students’ CBM scores. Previous studies indicated that the normative increase in mathematics CBM scores, defined as the weekly growth rate, for general students in Taiwan was 1.2 (Tsuei, 2008). The trend line (in red) represented the actual performance of the child. The actual performance was determined by a linear growth function, which was calculated on the basis of the slope through a least-squares regression analysis from the CBM scores.

Table 3 shows the growth rates of each subject on the mathematics CBM in each semester. All subjects showed positive growth rates on the math CBM in both semesters. In the first semester, the CBM growth rates of subjects A, B, and C (1.21, 2.87, and 1.22, respectively) were equal to or higher than the normative growth rate. All subjects showed higher than normative growth rates in the second semester. Subjects A, C, and D showed increased growth.
rates (1.54, 1.92, and 1.43, respectively) in the second semester. Although subject B showed a slightly lower growth rate in the second semester compared to the first semester, he still maintained the highest growth rate among all subjects. Overall, students with LD showed comparable growth rates to those of the general students at the end of second semester. In sum, all of the subjects showed positive math CBM learning in the face-to-face synchronous G-Math peer tutoring environment.

Figure 4. Student A’s CBM graphs: (1) first and (2) second semester

Figure 5. Student B’s CBM graphs: (1) first and (2) second semester

Figure 6. Student C’s CBM graphs: (1) first and (2) second semester

Figure 7. Student D’s CBM graphs: (1) first and (2) second semester
Table 3. Growth rates of students on mathematics CBM

<table>
<thead>
<tr>
<th>Subject</th>
<th>First semester</th>
<th>Second semester</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Growth rate</td>
<td>Linear growth function</td>
</tr>
<tr>
<td></td>
<td>Score/day (week)</td>
<td>Y = 14.38 + 0.17X</td>
</tr>
<tr>
<td>A</td>
<td>0.17 (1.21)</td>
<td>Y = -1.04 + 0.41X</td>
</tr>
<tr>
<td>B</td>
<td>0.41 (2.87)</td>
<td>Y = 18.46 + 0.17X</td>
</tr>
<tr>
<td>C</td>
<td>0.17 (1.22)</td>
<td>Y = 12.80 + 0.03X</td>
</tr>
</tbody>
</table>

Analysis of online behaviour in the G-Math peer tutoring system

Average time spent to solve a math problem online

Fluency of basic facts is essential for primary education students to master higher-order skills because these facts serve as a foundation for mathematical applications (Coddin, Chan-Iannetta, Palmer, & Lukito, 2009). The average time taken to solve math problems across problem type was analyzed by dividing the year into six phases (by months during each semester). To compare the average time that it took students with LD to solve a math problem online with that of general students, one general class of 32 students used the G-Math peer tutoring system once a month during the semester.

Figure 8. Average time spent solving a math problem: (A) computational problem-type (B) conceptual problem-type (C) application problem-type

Students with LD spent an average of 276.46 and 374.77 seconds on computation-type questions in the first and second semesters, respectively. Therefore, students needed more time to solve computational problems in the second than in the first semester. Students in the general classroom solved a computational math problem performed nearly twice as quickly as students with LD (Figure 8-A).

Figure 8-B compares students with LD and general students for conceptual-type questions. Students with LD gradually solved conceptual math problems faster in the second semester (M = 172.03 s) compared to the first semester (M = 183.72 s). Moreover, their performances gradually became closer to those of the general students.
The average time that students with LD took to solve application-type questions was 330.38 and 466.47 seconds in the first and second semester, respectively (Figure 8-C). Students with LD needed more time to solve application problems during the second than during the first semester. Although students with LD performed comparably to the general students in terms of time to perform application problems in the first semester, they needed more than twice the time to solve an application problem than general students in the second semester.

Use of the virtual manipulatives during the online peer tutoring activities

The teacher assigned different learning units for children to perform peer tutoring activities in every phase. The average number of object manipulations in the learning unit performed from the log files was calculated for each learning unit. For most categories of objects, students used more math objects to solve math problems during the second semester than during the first semester (Table 4, 5). Computation and integer symbols were the most frequently used math objects in both semesters. Geometry objects were used less often than other categories of objects. Fraction symbols (M = 16.61) and lengths and areas (M = 17.64) were used more in the second than in the first semester. Generally, students used the math objects to solve math problems according to the content of the learning unit, i.e., they used more length and area objects (M = 36) in the “Perimeter and Area” unit than in other units.

To explore the effects of math object manipulations on students’ mathematics, Pearson correlation analysis was adopted. The results indicated that there was a significant correlation between number of math object manipulations and students’ scores on the conceptual, computational problem-type and average scores of CBM in the second semester (r = .98, p < .05; r = .97, p < .05; r = .99, p < .01). Apparently, students used more math objects in the peer tutoring gained more scores on CBM.

<p>| Table 4. Average number of object manipulations in first semester |
|--------------------|----------------|---------------|----------------|---------------|---------------|----------------|</p>
<table>
<thead>
<tr>
<th>Category of Objects</th>
<th>Learning units</th>
<th>Multiplication</th>
<th>Division</th>
<th>Add and Subtract</th>
<th>Add and Subtract</th>
<th>Decimals</th>
<th>Fractions and Decimals</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Computation symbols</td>
<td>Integer</td>
<td>7.80</td>
<td>37.40</td>
<td>17.70</td>
<td>30.00</td>
<td>22.00</td>
<td>22.97</td>
<td></td>
</tr>
<tr>
<td>2. Integer symbols</td>
<td></td>
<td>28.00</td>
<td>50.80</td>
<td>18.00</td>
<td>5.50</td>
<td>7.00</td>
<td>21.86</td>
<td></td>
</tr>
<tr>
<td>3. Decimal symbols</td>
<td>Add and Subtract</td>
<td>23.60</td>
<td>18.80</td>
<td>6.33</td>
<td>20.00</td>
<td>1.00</td>
<td>13.95</td>
<td></td>
</tr>
<tr>
<td>4. Fraction symbols</td>
<td></td>
<td>8.80</td>
<td>20.40</td>
<td>2.67</td>
<td>5.50</td>
<td>9.00</td>
<td>9.27</td>
<td></td>
</tr>
<tr>
<td>5. General images</td>
<td></td>
<td>7.60</td>
<td>19.20</td>
<td>2.00</td>
<td>1.50</td>
<td>1.00</td>
<td>6.26</td>
<td></td>
</tr>
<tr>
<td>6. Lengths and areas</td>
<td></td>
<td>14.20</td>
<td>16.80</td>
<td>2.67</td>
<td>5.00</td>
<td>2.00</td>
<td>8.13</td>
<td></td>
</tr>
<tr>
<td>7. Weight and volume</td>
<td></td>
<td>8.80</td>
<td>15.00</td>
<td>1.67</td>
<td>0.00</td>
<td>2.00</td>
<td>5.49</td>
<td></td>
</tr>
<tr>
<td>8. Time and money</td>
<td></td>
<td>6.80</td>
<td>35.60</td>
<td>2.33</td>
<td>1.00</td>
<td>16.00</td>
<td>12.35</td>
<td></td>
</tr>
<tr>
<td>9. Protractor</td>
<td></td>
<td>3.60</td>
<td>17.60</td>
<td>2.33</td>
<td>4.00</td>
<td>5.00</td>
<td>6.51</td>
<td></td>
</tr>
<tr>
<td>10. Data, chart, graphs</td>
<td></td>
<td>3.20</td>
<td>3.80</td>
<td>5.00</td>
<td>3.00</td>
<td>1.00</td>
<td>3.20</td>
<td></td>
</tr>
<tr>
<td>11. 2D geometry</td>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>2.00</td>
<td>3.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>12. 3D geometry</td>
<td></td>
<td>0.00</td>
<td>0.80</td>
<td>0.00</td>
<td>0.50</td>
<td>0.00</td>
<td>0.26</td>
<td></td>
</tr>
</tbody>
</table>

Table 5. Average number of object manipulations in second semester

<table>
<thead>
<tr>
<th>Category of Objects</th>
<th>Learning units</th>
<th>Perimeter and Area</th>
<th>Fractions</th>
<th>Integer Multiplication</th>
<th>Time</th>
<th>Decimals</th>
<th>Integer Division</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Computation symbols</td>
<td>Integer</td>
<td>90.00</td>
<td>47.00</td>
<td>16.50</td>
<td>29.00</td>
<td>23.00</td>
<td>19.80</td>
<td>37.55</td>
</tr>
<tr>
<td>2. Integer symbols</td>
<td></td>
<td>48.00</td>
<td>43.50</td>
<td>34.00</td>
<td>63.00</td>
<td>67.70</td>
<td>105.00</td>
<td>60.19</td>
</tr>
<tr>
<td>3. Decimal symbols</td>
<td>Add and Subtract</td>
<td>8.00</td>
<td>4.00</td>
<td>21.00</td>
<td>23.00</td>
<td>4.33</td>
<td>28.20</td>
<td>14.76</td>
</tr>
<tr>
<td>4. Fraction symbols</td>
<td></td>
<td>22.00</td>
<td>8.00</td>
<td>2.50</td>
<td>8.00</td>
<td>28.30</td>
<td>30.80</td>
<td>16.61</td>
</tr>
<tr>
<td>5. General images</td>
<td></td>
<td>3.00</td>
<td>6.00</td>
<td>1.50</td>
<td>6.00</td>
<td>9.33</td>
<td>17.40</td>
<td>7.21</td>
</tr>
<tr>
<td>6. Lengths and areas</td>
<td></td>
<td>36.00</td>
<td>9.50</td>
<td>5.50</td>
<td>18.00</td>
<td>19.70</td>
<td>17.20</td>
<td>17.64</td>
</tr>
<tr>
<td>7. Weight and volume</td>
<td></td>
<td>8.00</td>
<td>7.00</td>
<td>22.50</td>
<td>8.00</td>
<td>22.30</td>
<td>15.40</td>
<td>13.87</td>
</tr>
<tr>
<td>8. Time and money</td>
<td></td>
<td>8.00</td>
<td>0.00</td>
<td>0.50</td>
<td>1.00</td>
<td>1.33</td>
<td>14.00</td>
<td>4.14</td>
</tr>
<tr>
<td>9. Protractor</td>
<td></td>
<td>10.00</td>
<td>1.00</td>
<td>3.50</td>
<td>4.00</td>
<td>8.67</td>
<td>14.80</td>
<td>6.99</td>
</tr>
<tr>
<td>10. Data, chart, graphs</td>
<td></td>
<td>9.00</td>
<td>0.00</td>
<td>0.00</td>
<td>12.00</td>
<td>11.00</td>
<td>9.20</td>
<td>6.87</td>
</tr>
<tr>
<td>11. 2D geometry</td>
<td></td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>18.00</td>
<td>0.00</td>
<td>1.60</td>
<td>3.43</td>
</tr>
<tr>
<td>12. 3D geometry</td>
<td></td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>36.80</td>
<td>6.30</td>
</tr>
</tbody>
</table>
Discussion and conclusions

The results supported the positive learning effects of G-Math on mathematics learning for students with LD and add to the body of literature in several ways. Participating students showed a positive improvement in terms of CBM scores and growth rates. Previous research illustrates the difficulty of effecting math achievement in the computer-mediated communication environment for elementary children (Meijden & Veenman, 2005). The present study demonstrated that the synchronous peer tutoring approach in the CSCL environment improved math learning for elementary students with LD, especially in conceptual math skills.

Unexpectedly, the online peer tutoring system did not help improve students’ computational scores with CBM, in contradiction with the previous research (Fuchs, Fuchs, Hamlett et al., 1997; Calhoon & Fuchs, 2003) showing the positive results of using peer tutoring to teach math computational skills in traditional classroom settings. This result may have been due to the need for students to manipulate more integer objects to complete the computational problems online, as compared to solving problems on a paper-and-pencil worksheet. It may be difficult to use the arithmetical format of computer tools for constructing a domain representation (Kolloffel, Eysink, & Jong, 2011). The average time that the children needed to solve a computation problem ($M_{(1st\, semester)} = 276.46\, s$, $M_{(2nd\, semester)} = 374.77\, s$) was larger than that needed to solve a conceptual problem ($M_{(1st\, semester)} = 172.03\, s$, $M_{(2nd\, semester)} = 183.72\, s$). This phenomenon showed that the online settings did not facilitate the fluency of students in solving computational problems. As Poncy, Skinner, and Jaspers (2007) stated, developing fluency with basic mathematics facts is critical for elementary students. The speed and accuracy with which math skills are completed may be valuable information for determining this fluency (Binder, 1996). The results of this study indicated that students with LD improved their fluency on conceptual problems, and that their fluency was approaching that of general students. However, their fluency performance on computation and application math problems has to be concerned when more complex problems were presented.

All of the subjects in the study showed improvement on application problems. These findings provided an extension to the accumulating research on face-to-face peer learning. Previous research indicated that students with LD exhibit difficulties in using higher-order math skills, such as cognitive and metacognitive problem-solving skills in application problems (Montague & Applegate, 1993; Calhoon & Fuchs, 2003). As Maccini and Hughes (2000) argued, manipulative instruction was effective in teaching word problem solving to students with LD. Our result indicated that students benefitted from the concrete and visualized participation through the tutor’s problem-solving processes. The visualized process facilitates students’ mathematics thinking. As Janssen, Erkens, Kanselaar and Jaspers stated (2007), the visualization of participation during online peer tutoring is one of the major contributions to successful CSCL.

The results of this study indicated that students gradually used more math objects in the peer tutoring activities. Integer numbers and computational symbols were the most frequently used objects. The results did show differences with regard to the inclination of students to use a representational tool. Charebout and Elen (2009) observed that representational tools integrated into a learning environment are often used inadequately or not at all by students. In the current study, the math objects used by students corresponded to the content of the learning units. Moreover, the results of the study indicated that students benefited from operating the math objects, especially for conceptual and computational math problems. This finding was consistent with previous work showing that students using arithmetical representative tools outperform students in the textual condition on their situational knowledge in mathematics (Kolloffel, Eysink, & Jong, 2011). Representations allow students to communicate mathematical approaches, arguments, and understanding to themselves and others. Therefore, the results of this study extend previous findings by indicating that providing manipulative materials in the synchronous online peer tutoring environment was effective in promoting the acquisition of mathematics skills.

We observed that both tutors and tutees provided comparable amounts of feedback sentences and emoticons to their partner. Students anticipated getting messages on the computer screen even though they seated side-by-side. When students played the tutee’s roles, they used question-asking sentences frequently. Usually, the tutor responded the tutee’s requests by repeatedly working on the same problems. Procedure information also included instances in which the tutor pointed out the specific operation to be performed (e.g., subtraction instead of addition) or specified which numbers to be used from the question sentences (e.g., Do you see the “36 pencils” in the questions?). However, the tutor’s elaborated explanations were still limited. As described in Bentz and Fuchs (1996), the peer tutoring training is important to encourage students providing more elaborated explanations. Therefore, more scaffolding
tools are needed in the future study. Moreover, students reported that they sometimes found it difficult to correct math errors to their partner. Although attempts were made to provide the correct answers in the G-math system, both LD students in the same pair may have had the same difficulties on certain types of math problems, which would have increased the difficulty in correcting math errors. Future works can develop the functions of generating the “adapted tutor” by applying Artificial Intelligence techniques when LD students have difficulties to teach each others. These results were in line with previous research supporting that peer tutoring is an effective pedagogical model in the CSCL context for elementary children (Meijden & Veenman, 2005; Tsuei, 2011, 2012) and extended previous CSCL research focused on peer tutoring to students with LD. Integration of the “game-like” and rewarding schema in the CSCL environment was effective to motivate students with LD to participate in remedial math activities. The teacher who participated in this study observed that students were highly motivated when they went to the resource classroom for playing the G-Math system. Students reported that they were eager to see the “upgraded” symbols of experience-value while they were online. The teacher also observed that students came to the resource rooms earlier because they wanted to meet their peer partners.

Although the findings from this study provide some promising information about the face-to-face synchronous peer tutoring system for elementary students with LD, there are several limitations. The sample size in the present study was small, and only curriculum-based measurements were collected. Further research may address these limitations by increasing the sample size and collecting additional qualitative data when appropriate. More qualitative data are also needed for analyzing the online interactions and their impact on problem solving skills in mathematics. Future research is needed to develop instructional strategies for enhancing arithmetic skills in the CSCL environment for students with special needs. One drawback of the G-Math system is that students only use mouse as input device to drag and drop math objects or draw corrections on tutee’s solutions. This feature limits the fluency computation skills of children. Future research in the CSCL field may also develop the peer tutoring system on the mobile tablet computer for elementary students to write the math answers on the screen and use more touch screen functions to promote peer tutoring skills.

References


Harrison, T. (2002). The development of a peer tutoring program to teach sight words to deaf elementary students. (Unpublished doctoral dissertation). The Ohio State University, Columbus, OH.


A Jigsaw-based Cooperative Learning Approach to Improve Learning Outcomes for Mobile Situated Learning

Yueh-Min Huang1*, Yi-Wen Liao2, Shu-Hsien Huang1 and Hsin-Chin Chen1

1Department of Engineering Science, National Cheng Kung University, Taiwan // 2Department of Information Management, Chia Nan University of Pharmacy and Science, Taiwan // huang@mail.ncku.edu.tw // pinkwen923@mail2000.com.tw // shuhsienhuang@gmail.com // luckyqq50@gmail.com

*Corresponding author

ABSTRACT

Most of the previous studies on mobile learning focused on using PDAs or cell phones to create situated learning contexts that enable students to learn anytime and anywhere, so as to increase their learning interest and motivation. However, some researchers have noted that the learning performance may be disappointing if such devices are used for situated learning without the support of appropriate learning strategies and instructor guidance. Therefore, this study proposed a jigsaw-based cooperative learning strategy with Google+ to support cooperative learning through social network services, and thus overcome the restrictions of traditional m-learning. In addition, Tablet PCs were used in this work as the learning device, as they have a better screen size than PDAs or cell phones, and this can strengthen the reading effect, providing a better learning environment. An experiment was conducted in the general curriculum “The ecological culture of Taiwan - the waters” to assess the effectiveness of the proposed approach. The experimental results show that this approach not only improved the students’ learning attitude, but also enhanced the effectiveness of learning.

Keywords

Cooperative learning, Mobile learning, Jigsaw-based strategy

Introduction

Ubiquitous mobile learning is now receiving increasing attention after that paid to distance and e-learning (Pownell & Bailey, 2001; Hwang, Tsai, & Yang, 2008; Wu, Hwang, Su & Huang, 2012). Quinn (2000) noted that m-learning is using mobile computing devices for learning, and as such it primarily relies on portable devices and a wireless network environment to enable users to access information (Seppala & Alamaki, 2003). Several studies have started to explore the benefits of handheld devices, such as smart phones and Tablet PCs, when applied to learning. For example, Huang, Lin, and Cheng (2010) found that they can increase learning motivation, and effectively enhance the resulting learning effects when combined with appropriate instructional strategies.

With the rapid evolution of information technology, many novel types of information technology equipments were used in learning environment (Huang, Liang & Chiu, 2013; Liu & Shen, 2011; Liu, Huang, Kinshuk & Wen, 2013). For example, handheld devices are continually being upgraded, and Tablet PCs have gradually started to replace more traditional m-learning devices, such as mobile phones and PDAs (Huang, Liang, Su & Chen, 2012). Since Tablet PCs support a broader range of multimedia audiovisual functions and have larger screens than either, PDAs or smart phones, they offer much better operational and learning environments than these other devices. Moreover, many learners already have a high degree of usage acceptance for Tablet PCs. Sommerich et al. (2007) suggested that another advantage of such devices is the diverse presentation of knowledge that they allow, with the digital information they present being easy to save, compress, browse, and carry. Ozoka et al. (2008) pointed out that the time and cost involved in printing traditional books is rather high, and thus Tablet PCs have the further benefit of lowering publishing costs. Moreover, Leeson (2006) indicated that using Tablet PCs to read articles is just as comfortable and convenient as using paper. Therefore, Tablet PCs have become the most promising devices for using in m-learning contexts.

In addition to the m-learning devices used, scholars have also noted the importance of constructing reliable learning activities based on real situations (Brown, Collins & Duguid, 1989; Wenger, 1997; Shih, Chuang & Hwang, 2010; Huang, Huang, & Lin, 2012). When people encounter problems they generally seek solutions to them, or ask more experienced people for help, and attempt to learn how to deal with the situation through a process of exploration (Brown, Collins & Duguid, 1989). For this reason, situated learning emphasizes that learning should take
place in real situations, as if only general abstract knowledge thus presented is being transmitted without the support of real situations, the knowledge is inert and difficult to transfer and assimilate. Meaningful learning should thus occur in real contexts, allowing learners to construct their own knowledge and solutions (Huang, Chiu, Liu & Chen, 2011). Some studies have also pointed out that if natural science courses carried out in an m-learning environment do not have suitable learning strategies and tools, then it may be too difficult for students to simultaneously deal with the challenges of the actual environment and also use the digital learning resources, and thus they may have disappointing learning outcomes (Hwang, Chu, Shih, Huang & Tsai, 2010), a view echoed in studies carried out in other m-learning contexts (Chen & Li, 2009; Liu, Peng, Wu & Lin, 2009). The question is thus how to combine real situational environments and digital learning resources to achieve the optimal learning effects in an m-learning environment.

While many recent studies focusing on m-learning and u-learning environments show that it is possible to achieve genuinely personalized learning using such approaches (Evans, 2008; Uzunboylu, Cavus & Ercag, 2009), few works explore the use of cooperative learning activities in such environments. The current work is thus aimed at introducing cooperative learning strategies to an m-learning environment, using group cooperation to increase interaction among peers, as this is expected to enhance students’ learning achievement and motivation and help cultivate important social techniques and concepts, such as respect for others, the use of clear expressions, rational communication, and the development of good interpersonal relationships.

Moreover, these positive effects can also continue after class, if the related learning activities are conducted on social networking sites. Stevens, Slavin and Farnish (1991) noted that in the process of cooperative learning, students experience group discussion, peer support and guidance, with an emphasis on personal responsibility and interdependence among group members, which not only all work to increase interactivity, but are also more efficient with regard to learning academic knowledge, solving problems, increasing learning motivation and cultivating positive learning attitudes. Cooperative learning has thus become an instructional strategy that is now quite broadly applied in many educational contexts.

Since jigsaw-based cooperative learning activities distribute tasks to all the members in the group, they not only increases student interaction, but can also get all the students to participate in the learning activities, thus enhancing the cooperative learning effects. This study thus uses jigsaw-based cooperative learning methods (Aronson & Patnoe, 1997) to improve student attitudes and the related learning effects by applying an m-learning cooperative approach. However, since the learning environment examined in this study is outdoors, and since the group members may have to spread out to engage in the activity, they are thus unable to engage real-time, face-to-face communication, lowering the effectiveness of jigsaw-based cooperative learning. In order to overcome this problem, this study adopted the Google+ social networking platform. This can be used to enable students and teachers to engage in both synchronous and asynchronous communication, as well as allowing the latter to better understand the learning status of each learner (Erkollar & Oberer, 2011). A pre-test, post-test, questionnaires and interviews were used in order to evaluate the feasibility of the learning activity designed in this work. Besides evaluating student learning effects, this study also explored the group cooperation and individual learning processes of students with high, medium, and low learning achievement. Finally, questionnaires and interviews were used to assess the suitability of this method for use a course on the ecology of Taiwan’s water regions, and student attitudes toward it.

**Jigsaw-based cooperative learning approach for mobile situated learning**

In the last few decades, there have been many studies on cooperative learning, examining the years between preschool and university, as well as many different academic subjects, and these demonstrate that the effects of using such approaches are far better than those associated with other instructional methods (Johnson & Johnson, 1990; Sharan, 1990; Slavin, 1995; Lin, Huang, & Cheng, 2010). Johnson and Johnson (1975) pointed out that successful cooperative learning requires two crucial elements: (1) the learning groups must promote the active learning of members through social interactions and discussions among; and (2) before the instruction, teachers should carefully design and arrange the course and provide the professional knowledge and guidance needed by the learners. Cooperative learning is basically a learning method that allows students to jointly achieve a given learning objective through the division of labor (Watson, 1991). In this type of learning process, group members have different responsibilities, and share each other’s learning accomplishments. Through social interaction, they can convey their understanding of certain concepts, assist each other, and jointly acquire new knowledge.
Jigsaw-based cooperative learning is an effective cooperative learning strategy, and has been shown to have good effects when applied to a broad range of academic subjects such as social studies, literature, and science (Slavin, 1995; Aronson & Patnoe, 1997). It not only increases the performance of students, but also promotes their communication abilities and interpersonal relationships (Slavin, 1989). Besides focusing on peer interaction and the cultivation of important abilities, such as critical thinking, problem-solving, and communication, jigsaw-based cooperative learning also helps students achieve a number of basic abilities, including independent thought, active exploration and research, clear expression, and team work. Jigsaw-based cooperative learning is carried out by assigning the same instructional materials to each group, but giving different parts to each group member. The members in each group who are responsible for the same parts of the material then form expert groups for joint study, and then go back to their home groups to teach what they have learned to their team members. This model is suited to activity units with more difficult learning materials, and can expand the range of cooperation from the group to the whole class, meaning that it conforms to the overall goal of cooperative learning. This instructional method has the following positive effects: (1) enabling students to effectively learn the instructional material; (2) enhancing their listening abilities and levels of engagement and empathy; and (3) raising the interdependence of students and joint learning.

The main characteristics of jigsaw-based cooperative learning are as follows (Aronson & Patnoe, 1997): (1) Students learn from the cooperation that occurs among members of the “home groups” and “expert groups,” so they can jointly research and share course-related information. (2) Heterogeneous groups: also known as home groups, which in this study are formed by students with high, medium, and low learning achievement, with three to six students in each group. In each home group, there are students who are responsible for the same sub-topic. These students then form a group to discuss the sub-topic, and this is known as an expert group, the members of which later share what they have learned with their home group. Jigsaw-based cooperative learning ensures that every group member can learn the focal concepts. Therefore, the main difference between this and other group cooperative learning methods is that in jigsaw-based cooperative learning, every team member has to be responsible for part of the instructional task, and thus everyone is involved in the activity.

Even though jigsaw-based activities can enhance the effects of cooperative learning, in real outdoor learning situations, group members may need to spread out in different places, potentially lowering the effectiveness of this approach. Therefore, it is necessary to have a learning platform that effectively connects the cooperation and discussion that occurs within groups. These cooperative learning tools or learning assistance mechanisms must provide real-time support, so that teachers and students can learn in a real context and construct their own knowledge, and teachers can grasp the learning conditions of each learner and thus be able to discuss any problems that arise, thus raising the positive effects of jigsaw-based cooperative learning.

Research questions

A Jigsaw-based cooperative learning approach is used for an m-learning activity in this study to explore the following research questions:
1. Can using a jigsaw-based cooperative learning approach produce better learning effects than traditional individual learning?
2. Can using a jigsaw-based cooperative learning approach enhance the learning effects of students with different learning abilities?
3. Can using a jigsaw-based cooperative learning approach enhance students’ cooperative learning attitudes?
4. Do students using a jigsaw-based cooperative learning approach think that engaging in cooperative learning is meaningful and interesting?

Experimental design

Due to changes in Taiwanese society, most students lack knowledge about the local ecology. Therefore, this study used a university general education course, “Ecological Culture in Taiwan - Water Regions,” with the aim of the learning activity being to increase students’ understanding of the various water regions in Taiwan through the observation of fish and the upstream, midstream, and downstream river ecologies. It was expected that this course could enable the students to learn about the differences among the various water regions in Taiwan, and their relation
to the learning objectives of the course. The learning environment was the National Museum of Marine Biology and Aquarium in southern Taiwan.

Learning environment

The National Museum of Marine Biology and Aquarium has a wireless network, allowing students to use Tablet PCs to go online and access the Google+ learning platform, and conduct discussions and engage in jigsaw-based cooperative learning as shown in Figure 1. The instructional content includes three main parts, examining the upstream, midstream, and downstream ecologies of rivers in Taiwan.

Participants

There are 63 students (31 males and 32 females) and 1 teacher (with 3 years of teaching experience) in the Information Management Department at a university in southern Taiwan volunteered to participate in this experiment. In order to ensure that the students were familiar with using Google+ on Tablet PC, an instructor spent fifty minutes teaching the students how to do this in a classroom on the day before the experiment. The students were then divided into experimental and control groups. The experimental group included 30 students (16 males and 14 females), who were divided into those with high, medium, and low learning achievement. The bottom 1/3 in the total academic grade rankings were the low-achievement students, the top 1/3 were the high-achievement ones, and the remaining 1/3 were the medium-achievement students. The results of one-way ANOVA showed the means of the different learning achievement groups were significantly different. In the experiment, the jigsaw-based cooperation activity was carried out using heterogeneous groups (Hooper and Hannafin, 1988), with three students in each group, one each from the high-, medium-, and low-achievers, for a total of 10 groups. There were 33 students in the control group, and they engaged in individual learning without group discussions.

Experimental procedure

Figure 2 shows the experimental flow of this study. Before the experiment began, the students needed to spend thirty minutes to complete a pre-test that examined their existing knowledge of the ecological environments of the waters
in Taiwan. In addition, a teacher spent twenty minutes introducing the procedure of the learning activity and the concept of jigsaw-based cooperation to the students in experimental group and control group, and they then spend ten minutes practicing this form of cooperation. In the actual learning activity, students in the experimental group used the Google+ learning platform to carry out jigsaw-based cooperation and engage in discussions in order to answer questions about the learning material. The platform contained three questions, and each student in each group was responsible for one of these. Students who were responsible for the same question in each group then gathered in expert groups on Google+, and discussed their opinions about the answer. After 40 minutes of discussion, they then went back to their home groups to share the information they had learned, and each group discussed all three answers for another 40 minutes. Finally, the teacher posed another question for cooperative discussions in the home groups, allowing group members to work together in order to answer it.

The students in the control group underwent a similar experience as the experimental group, and also used the Google+ learning platform to answer questions, but did not engage in any group discussions, instead working alone throughout the activity. During the activity, the teacher supervised the students’ learning conditions, and intervened if they needed some help. A pre-test was conducted before the experiment, and a post-test was conducted after it. The pre- and post-test scores were used to compare the students’ learning accomplishments with regard to the learning materials. In addition, a questionnaire survey and interviews were used to find the students’ views toward this activity, and suggest some avenues for further research. Figure 3(a) shows the students in the experimental group
engaging in real-time discussions while they observing the different ecological areas and Figure 3(b) shows a student working alone in the control group.

Google+ learning environment

Tablet PCs were used as the learning device, and Google+ as the learning platform. The functions of Google+ include location awareness, as well as the ability to post photographs, videos, and hyperlinks. Students were able to use instant messaging and the team videoconferencing function to engage in real-time discussions while observing
the different ecological areas. In Figure 4, a group of students in the experimental group are discussing the questions with other students or the teacher, and uploading an image of a fish to share with each other in the Google+ learning environment. In Figure 5, the students in control group are answering questions by directly observing ecological environment, without the aid of group discussions.

Measurement tools

In order to access the students’ knowledge of the ecological environment of the water regions in Taiwan before and after the intervention, pre- and post-tests were conducted. The pre-test consisted of 20 multiple choice questions, and each question had four possible answers, with each question worth five points, for a total score of 100. The post-test consisted of 14 multiple choice questions and three fill-in-the blank questions, for a total score of 100. The teachers who designed the pre- and the post-tests both teach natural science courses, and the tests were also evaluated by three professionals working in the field of natural science to ensure their validity.

In order to verify the internal consistencies of the pre- and post-tests, the questions were analyzed using the Cronbach’s α reliability coefficient. The pre-test result was 0.753, and the post-test result was 0.786, indicating that these tests had good internal consistency and reliability. In addition, questionnaire was used to assess the students’ learning interest and attitudes, with the items scored on a seven-point Likert scale. The questionnaire contained 15 items, including ones on the local cultural course, information technology innovation, perceived ease of use, perceived usefulness, perceived fun, usage motivation and attitudes, satisfaction, self-perceived usage effects, usage intention, continued usage intention, efficacy, individual and group cooperation satisfaction. The Cronbach’s α values for the experimental and control groups were 0.89 and 0.84, respectively, indicating that the questionnaire was reliable.

Experimental results

In order to evaluate the activity used in this study, SPSS for Windows was used to analyze the results of the pre- and post-tests, the scores of the high, medium, and low learning achievement students, and the questionnaire and interviews. Besides evaluating the students’ learning effects, this analysis also explored the cooperative and individual learning processes of the students with high, medium, and low learning achievement. The questionnaire and interviews were used to assess the suitability of the intervention, and how satisfied the students were with it.

Learning achievement

The experimental group engaged in cooperative learning while the control group engaged in individual learning. Both groups used the Google+ platform to carry out situated learning, although only the former engaged in group discussions, and the learning achievement test examined whether there were any significant differences between the outcomes for the two groups. The pre-test scores were used as the covariate for analysis of covariance (ANCOVA) to avoid any interaction effects from the pre-test on the students’ learning outcomes. The statistical analysis used the post-test scores as the dependent variable, the instructional method as the independent variable, the pre-test scores as the covariant, and ANCOVA to explore whether the experiment and control groups saw any significant differences in their learning effects. The test of the homogeneity of the regression coefficients showed that $F=.363 (p>.05)$, which did not reach a level of significance, and thus the null hypothesis was accepted. In other words, the two groups had the same slope in their regression lines, which conforms to the hypothesis of the homogeneity of the regression coefficients of the groups, and this ANCOVA could be used.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Adjusted Mean</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post-test</td>
<td>Experimental</td>
<td>30</td>
<td>75.13</td>
<td>6.89</td>
<td>76.671</td>
<td>4.76*</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>33</td>
<td>56.15</td>
<td>7.52</td>
<td>56.762</td>
<td></td>
</tr>
</tbody>
</table>

Note. SD: Standard deviation; *$p < .05$
In ANCOVA, the post-test scores were used to explore whether there were any significant differences between the instructional methods. The results showed that $F=4.767\ (p<.05)$, which reached a level of significance (Table 1). In other words, the two post-test scores of both groups showed significant differences due to the different instructional methods. After excluding the influence of the pre-test, the marginal means of the experimental group’s post-test was 76.671, higher than the control group’s 56.762, and the learning achievement post-test scores of the two groups also showed a significant difference. The results indicated that using jigsaw-based cooperative learning for situated ecological instruction is better than individual independent learning. A possible reason for this is that the jigsaw-based group cooperative learning model distributes various tasks to the group members, and requires some students to complete their part of the task before returning to the home group to share what they have learned. Due to the positive pressure of peer competition, group members did their best to help their partners. Therefore, the effects of group learning are better than those of individual learning in the situated learning environment.

**Effects of high, medium, and low learning achievement students**

In order to understand whether the level of learning achievement would affect the relationships seen in team cooperation, this study divided the students into those with high, medium, and low learning achievement. The bottom 1/3 in the total academic grade rankings were the low-achievement students, the top 1/3 were the high-achievement ones, and the remaining 1/3 were the medium-achievement ones. One-way ANOVA was conducted to determine whether the mean of the different learning achievement groups were significantly different, and the results showed significant differences with regard to the pre-test, $F(2,63)=12.04,\ p<.01$. The results of the pairwise comparisons with Tukey post hoc tests indicated that the low-achievement group performed significantly worse and high-achievement one significantly better. Moreover, the learning achievements of the low-, medium-, and high-achievement students were compared based on their post-test grades, and these were examined using $t$-test analysis to assess the differences.

**Low-achievement students**

In the analysis of the subjects’ learning achievement, there were 21 low-achievement students, 10 of whom were in the experimental group and used Google+ for group learning, and 11 of whom were in the control group and used Google+ for individual learning. An independent sample $t$-test was conducted, with the results shown in Table 2, and these indicate that using jigsaw-based group cooperation to engage in situated learning lead to significantly better learning outcomes than undertaking individual situated learning ($p<.05$). In other words, low-achievement students had greater accomplishments with jigsaw-based group cooperative situated learning than individual situated learning, most likely because of the assistance provided by other group members, as well as the active roles that they were forced to take as experts.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>10</td>
<td>70.89</td>
<td>5.31</td>
<td>2.86</td>
<td>.01*</td>
</tr>
<tr>
<td>Control</td>
<td>11</td>
<td>50.37</td>
<td>12.73</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. SD: Standard deviation; *$p<.05$*

**Medium-achievement students**

Twenty-one students were classed as medium-achievement one, 10 of whom were in the experimental group and used Google+ for group learning, while 11 were in the control group used Google+ for individual learning. An independent sample $t$-test was conducted, and the results in Table 3 indicate that using jigsaw-based group cooperation to engage in situated learning lead to better learning outcomes for individual situated learning ($p<.05$). This shows that the distribution of tasks in the jigsaw-based groups also helped the medium-achievement students, by letting them receive assistance from the high-achievement students, and keeping them engaged in the task.
Table 3. Medium-achievement students’ independent sample t-test results

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>10</td>
<td>74.89</td>
<td>4.93</td>
<td>3.77</td>
<td>.03</td>
</tr>
<tr>
<td>Control</td>
<td>11</td>
<td>60.74</td>
<td>14.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. SD: Standard deviation; *p < .05

High-achievement students

There were 21 high-achievement students, 10 in the experimental group who used Google+ in group learning, and 11 in the control group who used Google+ for individual learning. An independent sample t-test was conducted, and the results are shown in Table 4, which reveal that using jigsaw-based group cooperation to engage in situated learning did not lead significantly better outcomes than individual situated learning (p>.05). This is perhaps because the high-achievement students had leadership roles in the group learning activity, and thus spent a lot of time helping their partners. For such students, individual situated learning may be more effective than group cooperation. In addition, in the questionnaires and interviews, most of the high-achievement students also stated that they preferred individual rather than group learning in situated instructional environments.

Table 4. High-achievement students’ independent sample t-test results

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>10</td>
<td>76.61</td>
<td>2.93</td>
<td>4.86</td>
<td>.17</td>
</tr>
<tr>
<td>Control</td>
<td>11</td>
<td>57.34</td>
<td>10.41</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. SD: Standard deviation; *p < .05

On the whole, for low-and medium-achievement students, using jigsaw-based group cooperative learning may lead to better outcomes than individual situated learning. The jigsaw-based group task division method can reduce the burden on low- and medium-achievement students, helping them to take part in the discussions, raising their learning interest, increasing learning motivation, and in turn elevating the learning effects. This approach can thus not only close the learning gaps that exist in class, but can also enhance instructional quality.

Questionnaires and Interviews

In order to better understand students’ attitudes towards the situated learning activity based on the ecological environment of water regions in Taiwan, a questionnaire survey was completed by 63 students in the experimental and control groups after the activity had been completed. A total of 60 valid questionnaires were returned. The questionnaire contained 15 items, which were measured using a seven-point Likert scale. The questionnaire validity values ranged from 0.79 to 0.73, and thus these were of reliable consistency, as shown in Table 5.

Table 5. Questionnaire results

<table>
<thead>
<tr>
<th>Questions</th>
<th>Groups</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I have confidence that I can complete the course by using the Google+</td>
<td>Experimental</td>
<td>5.50</td>
<td>1.30</td>
</tr>
<tr>
<td>learning platform.</td>
<td>Control</td>
<td>4.86</td>
<td>1.71</td>
</tr>
<tr>
<td>2. I am interested in the information that the Google+ learning platform</td>
<td>Experimental</td>
<td>5.13</td>
<td>1.75</td>
</tr>
<tr>
<td>provides.</td>
<td>Control</td>
<td>5.10</td>
<td>1.47</td>
</tr>
<tr>
<td>3. I feel the Google+ learning platform can raise my curiosity.</td>
<td>Experimental</td>
<td>4.83</td>
<td>1.61</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>5.13</td>
<td>1.67</td>
</tr>
<tr>
<td>4. I would like to recommend the Google+ learning platform to other people.</td>
<td>Experimental</td>
<td>5.43</td>
<td>1.27</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>5.39</td>
<td>1.69</td>
</tr>
<tr>
<td>5. The Google+ learning platform helps me learn knowledge easily.</td>
<td>Experimental</td>
<td>5.93</td>
<td>1.52</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>5.46</td>
<td>1.08</td>
</tr>
<tr>
<td>6. I like to use the Google+ learning platform to carry out cooperative</td>
<td>Experimental</td>
<td>5.20</td>
<td>1.25</td>
</tr>
<tr>
<td>(individual) learning.</td>
<td>Control</td>
<td>4.76</td>
<td>1.88</td>
</tr>
<tr>
<td>7. I am satisfied with my performance in the cooperative (individual)</td>
<td>Experimental</td>
<td>5.13</td>
<td>1.90</td>
</tr>
<tr>
<td>learning activity.</td>
<td>Control</td>
<td>4.56</td>
<td>1.92</td>
</tr>
<tr>
<td>8. I like this method of cooperative (individual) learning.</td>
<td>Experimental</td>
<td>5.90</td>
<td>0.92</td>
</tr>
</tbody>
</table>

136
The results of the questionnaire survey showed that the students were satisfied with the use of Google+, and that the experimental group was more satisfied than the control group. The results also showed that the situated learning environment allowed for better individual learning, and that most students enjoyed learning with the jigsaw-based group cooperation method. In order to gain a better understanding of student attitudes toward this intervention, members of the experimental and control groups were randomly sampled for in-depth interviews, with one low-, medium-, and high-achievement students taken from each group, for a total of six students.

The results showed that in the experimental group, the low-achievement students liked the jigsaw-based cooperative learning the most. However, while the high-achievement student stated that this instructional method could increase interaction among peers, they also said that because they had to help their partners, they could not learn at their own pace, and thus they preferred individual learning in the situated instructional environment.

In addition, the high- and medium-achievement students in the control group both stated that in this type of instructional environment they preferred individual learning so that they could move at their own speed, with more time to observe the details of the ecological system being examined, although they noted that they felt a little bored as they were not able to share their ideas with classmates. The low-achievement students stated that since they usually learn slowly on their own, they hoped that discussions with their classmates could make this processes easier and faster, so that they would not have to ask the teacher directly when they encountered a problem with the activity. Since the teachers working with this class could not always answer questions when needed this raised the learning anxiety of the low-achievement students.

The results of both the quantitative analysis and interviews revealed that the intervention had more positive effects on the low- and medium-achievement students. In future situated instructional environments, instructors should consider using the jigsaw-based group cooperation method, as this can not only increase peer interaction, but also allow most students to improve based on positive peer competition, bringing the class closer together and enhancing the instructional quality of the teaching, as well as the resulting learning outcomes.

**Discussion**

This study used Tablet PCs as learning devices, as these have better screens than smart phones or PDAs, thus providing a better learning environment, especially for mobile collaborative learning. In addition, this study combined jigsaw-based cooperative learning with the Google+ social networking platform for an experiment involving an activity looking at the varied ecologies in the different water regions of Taiwan in order to evaluate the efficacy of this approach. This method was used alongside traditional individual learning to assess student performance in both contexts. The experimental results demonstrated that using the jigsaw-based mobile cooperative learning method could lead to better learning outcomes and attitudes in the experimental group than in the control one.
The use of effective cooperative learning tools and strategies can overcome some of the inadequacies of traditional learning approaches, and the method used in this study was able to enhance the post-test performance of the students in the experimental group. The results of the questionnaire also showed that the students in the experimental group felt that using the jigsaw-based mobile cooperative learning method could improve their learning performance. However, it should be noted that the high-achievement students preferred to complete their studies independently, although the medium- and low-achievement ones liked to work using positive peer competition and cooperation.

The results of this study show that jigsaw-based cooperative learning combined with the Google+ social networking platform did help the students acquire, discuss and organize their knowledge. Most recent studies of m-learning and u-learning have noted that in a situated learning environment, students can adjust their learning progress according to their own pace (Uzunboylu et al., 2009; Huang, Lin, & Cheng, 2010; Evans, 2008; Hwang, Chu, Shih, Huang, & Tsai, 2010; Huang, Chiu, Liu & Chen, 2011), and the experimental results and interviews carried out in this study showed that the learning environment used in this work was able to enhance learning outcomes, although these still differed for students with different levels of learning achievement. However, overall the results of this study suggest that jigsaw-based cooperative learning should be used in situated instructional environments, as this will increase interaction among peers, and the positive competition that arises from this can also enhance learning intention.

Conclusion

With the rapid development of technology, instructional environments have also become more diverse, from traditional paper and classroom-based approaches to online learning platforms, e-learning, m-learning, and u-learning, and the instructional focus has thus turned from the teachers to students. However, these new instructional contexts still need effective instructional strategies and learning tools if they are to produce good learning effects and greater satisfaction among learners. This study used Tablet PCs, Google+, and jigsaw-based cooperative learning in a university general education course called “Ecological Culture in Taiwan - Water Regions,” with the learning environment being the National Museum of Marine Biology and Aquarium in southern Taiwan.

This study found that for both individual and group learning, the students felt that this type of instructional activity could raise their learning interest and intention, leading to better learning outcomes. A more in-depth exploration of the learning achievement results revealed that attitudes toward the activity differed based on the students’ level of learning achievement. Students with high learning achievement stated that since the learning activity was connected to a real situation, they could explore various topics, and thus they preferred independent learning. However, the students with medium and low learning achievement felt that the jigsaw-based group cooperation method was better than individual learning. The jigsaw-based approach provides tasks for each group member, and even though some of the low-achievement students stated that these were difficult, they were able to discuss their problems with their partners, which resulted in a higher sense of accomplishment and participation. The medium-achievement students stated that they enjoyed working with the other group members who had higher or lower achievement, and that the positive competition and interaction that occurred among them could enhance their learning intentions. Overall, the students stated that using a jigsaw-based cooperative learning approach with Google+ to engage in cooperative learning was both meaningful and interesting. When teaching students with mixed abilities, it is thus suggested that the jigsaw-based cooperative learning model should be applied with the aid of Google+ in situated instructional environments, as this can increase the effectiveness of learning and enhance students’ cooperative learning attitudes.

Due to limitations of time, manpower, and resources, this study only focused on university students. In addition, the number of samples and computer literacy issues were not considered, and these are limitations of this study, which may limit the general ability of the results. If possible, future studies can use a greater number and variety of subjects, and consider aspects such as gender, different majors, and different levels of computer literacy. Meanwhile, other m-learning environments should be considered in follow-up research. Such studies would be able to engage in more detailed analysis of the students’ learning intentions and acceptance of this approach.

Acknowledgments

The authors would like to thank the National Science Council of the Republic of China for financially supporting this research under Contract No. NSC 101-2511-S-041-001-MY3, NSC 100-2511-S-006-015-MY3, NSC 100-2511-S-006-014-MY3 and NSC 101-3113-P-006-023-.
References


The Integration of Concept Mapping in a Dynamic Assessment Model for Teaching and Learning Accounting

David Tawei Ku1, Ju-ling Shih2* and Su-Huan Hung2

1Department of Educational Technology, Tamkang University, New Taipei City 25137, Taiwan // 2Department of Information and Learning Technology, National University of Tainan, Tainan City 70005, Taiwan // dtk@mail.tku.edu.tw // juling@mail.nutn.edu.tw // abby5917@yahoo.com.tw

*Corresponding author

(Submitted September 13, 2012; Revised December 19, 2012; Accepted March 6, 2013)

ABSTRACT

The purpose of this study is to integrate the idea of concept map into dynamic assessment model for accounting education in vocational high school. By using the expert concept map and the objectives as the main reference to generate assessment questions for diagnosis purposes, students should be informed the shortcoming of learning and receive proper guidance to improve learning. A quasi-experimental single group pre-posttest design was conducted for this study. Subjects were given a pretest before treatment and posttest after treatment. In order to truly compare the performance between the treatments, the final grades of previous semester course were used to distinguish three different (high, medium, and low scores) student groups. The results show that after the intervention of computerized dynamic assessment system, posttest scores were significantly higher than pretest scores in all three groups. In addition, regardless of learning styles, when students were willing to commit time to follow the guidance, they could result in a good learning progress. Moreover, students with high degrees of completion performed significantly better than low degrees of completion students in posttest-pretest progress. Recommendations for using dynamic assessment system in accounting learning are also discussed.

Keywords

Concept map, Dynamic assessment model, Computerized dynamic assessment system, Accounting

Introduction

Low learning willingness and sluggish understanding typically result in low academic achievement because students fail to thoroughly and effectively learn. Without appropriate after-school learning programs or remedial strategies, students can easily fall behind and lose interest in learning. To diagnose students’ learning difficulties, provide students with remedial strategies for learning problems, and enable students to develop their learning potential, this study implements a dynamic assessment model in the instructional process.

Many studies have highlighted that learning assessments effectively promote the improvement of teaching and learning (e.g., Airasian, 1989, 1986; Linn & Gronlund, 1995; Linn & Miller, 2005). However, traditional static learning assessments are limited in their applicability for evaluating students' learning achievements, identifying their learning potential, and providing precise and useful solutions to learning difficulties. Vygotsky (1978) proposed the concept of the zone of proximal development (ZPD), which is defined as “the distance between the actual developmental level, as determined by independent problem solving, and the level of potential development, as determined through problem solving under adult guidance, or in collaboration with more capable peers” (p. 86). This highlights the feasibility of probing and exploring individual learning potential through guidance. Feuerstein (1979) found that traditional assessments emphasize students’ learning failures instead of their learning process and learning potential; thus, he developed a dynamic assessment tool for implementing corresponding improvements. Through teacher guidance, dynamic assessments can diagnose students’ learning difficulties, provide intermediary assistance, and create meaningful interactions that stimulate students’ learning potential.

Based on the concept of dynamic assessments, Vye, Burns, Delcos, and Bransford (1985, 1987), and other scholars advocated the development of a continuous dynamic assessment model that comprised the following two main cores: the standardized graduated prompting assessment and non-standardized meditational assessment models. Research has also indicated that dynamic assessments can effectively promote individual learning potential and achievement, enhance the relevance of learning content, provide intermediary teaching assistance, and develop a more efficient remedial teaching system (Xu, 2004).
However, dynamic assessments do not always provide customized learning remedies. Typically, they identify learning difficulties or failures without offering solutions to improve student learning. To develop a more effective remedial teaching and learning system, this study employs a concept map established by experts (teachers) to structure a curriculum that integrates learning content and provides the required teaching and learning contexts.

Novak (1991) proposed a new concept of learning based on prior acquisition. Functionally, concept maps can be regarded as scaffolding that supports learning and assessments. During teaching activities, instructors serve as experts who understand the in-depth conceptual structure of the learning content. If instructors provide a concept map of the learning content, learners can understand the conceptual structure of experts and follow a guided learning path. Concept maps can also be used for learning assessments. According to the stage and difficulty level of student mistakes, a concept map can identify the area of mistakes and provide timely remedial instruction.

In recent years, concept maps have been gradually computerized. Using a concept map created for a specific subject, relevant systems can identify learning difficulties and student advantages and provide appropriate individual learning guidance to enhance academic performance (Hwang, 2003; Chang, Sung, Chang & Lin, 2005; Liu, Chen & Chang, 2010). Studies have shown that students who construct personal concept maps are able to identify misconceptions. Accordingly, computer systems and teaching instructors can provide students with timely support and guidance based on these misconceptions.

Concept maps of teaching content constructed by experts should be more comprehensive and organized than those constructed by students. Thus, instructors can also adopt expert concept maps when designing learning instruction and activities to further benefit students.

This study used the Accounting Entry section of a second-year vocational high school Accounting program as the primary subject. Employing an expert concept map to establish the relevance of learning concepts for intermediary support of the teaching context, a computerized dynamic assessment system was developed. By referencing an expert concept map and teaching objectives when developing assessment questions, students can exploit the adaptive functions to explore their learning difficulties and receive appropriate guidance to improve learning performance.

**Literature review**

**Dynamic assessment**

Budoff (1974) proposed the learning potential assessment model (test – train – test) to assess the abilities of educable mentally handicapped children. Based on cognitive development theory proposed by Piaget (1950), Paour (1992) suggested that assessments should incorporate logical thinking instructions to evaluate changes in children’s cognitive structure. Feuerstein (1979) further highlighted that mediated learning is a necessary condition for cognitive development, and proposed the Learning Potential Assessment Device (LPAD; pretest – mediate – posttest).

Based on the ZPD established by Vygotsky (1978), Campione and Brown (1987) proposed the graduated prompting assessment model (pretest – learning – transfer – posttest) as an alternative to the non-standardized intervention developed by Feuerstein (1979), using a standardized progressive teaching support as an intermediary. Embretson (1987) emphasized the variability of cognition and developed a psychometrics-oriented dynamic assessment model (psychometric approach), which employs spatial reasoning tests as training materials and uses “pretest – train – posttest” procedures to assess learning results.

Burns, Vye, and Bransford (1987) indicated that learning is a continuous direct experience. Burns and colleagues integrated the LPAD developed by Feuerstein (1979) and the graduated prompting assessment model established by Campione and Brown (1987) to construct the continuum of assessment model. The experimental results showed that intermediary teaching can better promote learning transfers compared to progressive prompting and enhance learner performance in dynamic assessments.

Ding (2002) stated that the current mainstream form for assessments is the standardized intervention “pretest – intervention – posttest” model, which is primarily based on the graduated prompting assessment model of Campione
and Brown (1987). Because the standardized assessment system design can be easily extended to various subjects, it is the most commonly used dynamic assessment model. Moreover, by combining the standardized and non-standardized intermediary teaching processes of multi-stage dynamic assessments, Xu (2004) found that continuous assessments can eliminate learning problems and enhance learners’ problem-solving abilities.

Therefore, this study adopts the continuum assessment model as the core system architecture for the Accounting Entry subject. By integrating a concept map, this system can investigate students’ learning difficulties and identify potential learning problems to facilitate learning and promote effective learning and teaching.

Concept mapping

Based on meaningful learning theory established by Ausubel (1963), Novak and Gowin (1984) developed the concept mapping method. In the book Learning How to Learn, they defined concept maps as a graphical metacognitive tool for learners to express their understanding of the knowledge theme.

Concept mapping involves using the appropriate link to connect one or more concepts, thereby expressing the complete meaning and relationship between the concepts. Concept maps can also provide a visualization of the hierarchy and structure of a system. Furthermore, teachers can use concept maps drawn by students to assess their understanding, logical ordering abilities, and other capabilities in relation to the particular subject or concept.

Since Novak and Gowin (1984) applied concept mapping to the field of science learning, it has been employed in other disciplines for teaching, learning, searching, and assessment. Concept maps have also been verified to be effective learning tools that promote positive learning attitudes and achievement (Allan, 2003; Nakhleh, 1994). Moreover, concept maps can be used as a learning diagnostic tool for remedial instruction and provide clear guidance for low achievers or solutions to learning difficulties (Tsai, Lin, & Yuan, 2001; Donald & Barbara, 2003).

Research of concept maps focuses on the following three main applications:
1. As a teaching method: Teachers can include concept maps in the teaching unit as a formative assessment or convergence tool to connect units.
2. As a tool for learning and teaching strategies: Concept maps benefit knowledge reconstruction and retention and can enable students to learn concepts and derivative meanings, thereby expanding the scope of a single learned point.
3. As a tool for assessing and identifying misconceptions: Concept maps are not only extremely powerful learning tools but also effective assessment tools. They enable students to confirm whether their knowledge of a concept is correct (Mintzes, Wandersee, & Novak, 2000). In addition, concept maps can be used to assess concept structuring and change processes, and are a good tool for identifying misconceptions and knowledge flaws.

Zeilik et al. (1997) applied concept maps to the subject of Astronomy and found that students experienced tremendous success in distinguishing misconceptions and learning concepts. Heather, Joel, and Richard (2003) used concept maps as a tool for assessing learning outcomes, where students could benefit from the entire knowledge base of a concept, instead of depending on the trivial fragments retained.

Computerized concept mapping

Donovan (1983) highlighted that concept mapping is the process of constructing meaning. The main processes are as follows: (a) Screen important concepts; (b) reflect the concept of chronological order; (c) use appropriate links to show the relationship between concepts; (d) cross-link the relationships between concept branches; and (e) provide specific examples that illustrate the concepts. Concept mapping is an effective teaching technique that enables students to build an organization and illustrate changes in cognitive structures. In addition, concept mapping serves as an important strategy for improving the concept of learning.

However, the traditional paper and pencil style of concept mapping is extremely inconvenient and a burden for instructors. With the rapid development of computer technology, the use of computerized concept maps to support teaching and learning has gradually increased. Computerized concept mapping systems can be adapted to individual differences and provide diverse teaching methods that attract learners’ attention. More importantly, the calculation
speed and analysis ability of a computerized system provides multidimensional organized information for diagnosis, remedial instruction, and learning.

Ausubel (1963) highlighted that meaningful learning is based on students’ prior knowledge and builds new knowledge. Traditional teaching methods involve teachers providing guidance and presenting appropriate materials. As experts, teachers design the structure of a curriculum and its activities, and students follow the design to learn gradually.

By presenting teachers’ instructional architecture as a concept map, this study develops a dynamic assessment system based on an expert concept map. When employed, this dynamic assessment system can identify incorrect items, determine the error type, and provide guidance by directing students to the appropriate content unit for review based on the expert concept map. Through the system architecture of an expert concept map, the assessment process follows a continuous learning model. According to the system responses, learners can enhance their learning processes and learning effectiveness.

Learning styles and profiles

Identifying learning styles to facilitate teaching and learning is a crucial issue in education. Keefe (1979) classified learning styles as “characteristic cognitive, affective, and psychological behaviors that serve as relatively stable indicators of how learners perceive, interact with, and respond to learning environments” (p. 4). Learning styles not only indicate the learning characteristics of learners, but can also imply possible disciplinary differences (Kolb, 1981). Interestingly, learners’ learning tendencies and characteristics can also be used to identify their learning style. That is, by analyzing learners’ learning processes, their learning problems and preferred learning styles can be identified. Traditional assessments only provide short-term and partial learning outcomes; however, learning profiles can provide instructors with crucial information such as learning paths and preferences.

As mentioned previously, concept mapping enables learners to accumulate knowledge by linking concepts logically or even chronologically (Donovan, 1983). According to Felder and Silverman (1988), sequential/global learning style dimensions can be well described by constructing concept maps. Sequential learners progress toward understanding in continual steps, whereas global learners holistically develop comprehension in large increments. By employing this system, the dynamic assessment model and expert concept map allow learners to identify learning difficulties. Student learning profiles also supply unique information that enables instructors to provide appropriate guidance to various individuals.

Research methods

System design

As the foundation of accounting, Accounting Entry includes the concepts of accounting classification, principle of duality, and double-entry bookkeeping. Any unclear concept will lead to an incomplete knowledge of Accounting Entry, and hinder successful learning of subsequent areas of accounting. Therefore, Accounting Entry is an appropriate subject for this study because a dynamic assessment system can facilitate the resolution of students’ learning difficulties. The system also reinforces adaptive remedial instruction, which enables students to develop a complete understanding of Accounting Entry and gain confidence in their ability to learn accounting.

The structure of the continuum dynamic assessment model is shown in Figure 1. Students can begin from any learning area. However, during the formal test, the students are instructed to complete each unit’s Accounting Entry area to pass the evaluation.

To clarify the system structure based on the hierarchical concept mapping design concept (Stice & Alvarez, 1987), the concepts of the content areas mentioned previously were gradually extended into an expert concept map, as shown in Figure 2. The oval represents the three conceptual areas of “accounting classification,” “principles of duality,” and “accounting entry.” In each area, the rectangle represents the sub-concept, and all concepts are connected to each other by various means. The arrows denote a linear connection between two concepts, and the
dotted lines indicate that the concept can be extended to various concept areas, of which the acronym PC in the figure represents “possible cause.” This expert concept map becomes the underlying content structure of the system, which can be regarded as the road map of the dynamic system. The system evaluates students’ understanding in each concept, shows their learning difficulties, and identifies potential learning problems when students fail to answer the questions. Moreover, the system offers adaptive assistance by bringing them back and forth to the specific concepts they should be familiar with until they complete the course.

Figure 1. The content structure of the dynamic assessment system for accounting

Figure 2. The expert concept map of the content of Accounting Entry

The system interface includes three display areas, from top to button: the main menu, sub-menu, and individual progress display (Figure 3).

1. Main menu: Includes three major (main content) sections, accounting classification, duality principles, and accounting entry. Students can access all assessment sections here, and the portfolio button takes students to check their current learning progress.

2. Sub-menu: This is the question-display page. When students select one of the major sections, five sub-concept buttons (assets, liabilities, owner equity, revenue, and loss) appear for students to readily access each unit.
3. Personal progress display: After students complete unit assessment, a personal progress display presents their status. Therefore, students can adjust their learning pace and develop effective learning.

Figure 3. Main user interface of the dynamic assessment system for accounting

Research design

The main subjects of this study were 34 accounting sophomores of a vocational high school in Taiwan. Integrating continuous dynamic assessments (the pretest – train – retest – train – retest…posttest model) and concept mapping, this study develops a computerized dynamic assessment system based on the content of the three main areas Accounting Classification, the Principles of Duality, and Accounting Entries.

Standard face-to-face teaching activities are still performed in the classroom; however, if required, the computerized dynamic assessment system is used to assess the learning results, and the recommend remedial strategies are implemented. Thus, the combination of remedial teaching and assessments achieve maximum effectiveness of the remedial program.

A quasi-experimental single-group pre-posttest design was employed for this study. The participants were administered a pretest before the intervention and a posttest after the intervention. Two tests were identical. Furthermore, to truly compare the performance before and after the intervention, grades for the accounting course from the previous semester were used to create three student groups, that is, a high-scoring group (top 29%, 10 students), low-scoring group (bottom 29%, 10 students), and medium-scoring group (42%, 14 students).

Research instruments

Achievement tests

To evaluate the reliability of the pretest, 30 senior accounting students were selected to complete the pre-stage achievement test, and the results were used to determine appropriate questions.

1. Scoring: A total of 16 questions were included in the pretest, where each question was worth 1 point, for a total of 16 points. The questions covered all learning modules of Accounting Entry.

2. Item analysis: According to the test results, the high- (11 students; top 37%) and low-scoring groups (11 students; bottom 37%) were separated, and the difficulty and discrimination of the items were determined using the following formula: The higher the \( p \) value, the higher the rate of difficulty; the higher the D value, the more accurate the discrimination rate.

\[
\text{Difficulty (P)} = \frac{P_H + P_L}{2}
\]

\[
\text{Discrimination (D)} = P_H - P_L
\]

\( P_H \) = correction percentage of the high-scoring group

\( P_L \) = correction percentage of the low-scoring group
According to Ebel (1979), a value over 0.3 indicates an acceptable discriminating ability. Huang and Jian (1991) highlighted that research should use the item analysis results to first select the items of high discrimination, followed by the items of moderate discrimination. Thus, the difficulty rate for the majority of the items will range between 0.35 and 0.65. Even when additional items, the overall difficulty rate should remain approximate to normal distribution.

With the exception of Items 7 and 16, all items possessed a discrimination rate above .30, and a difficulty rate between .23 and .77, which indicates that the items possessed good discrimination and appropriate difficulty levels. Item 7 (.47) was a moderately difficult question, with an acceptable discrimination rate of .27. Therefore, the subject matter experts recommended retaining Item 7. Item 16 was rated a high-difficulty question (.26); however, the discrimination rate was acceptable. According to the two subject matter experts, the concept of Item 16 was extremely important for the subject scope. Therefore, Item 16 was retained and modified to reduce the difficulty.

In addition, the overall Cronbach’s alpha value for the pretest was .788, which indicates good reliability.

**Data collection**

**Learning performance**

According to the accounting grades from the previous semester, the students were divided into high- (top 29%), medium- (mid 42%), and low-scoring (last 29%) groups. A paired t-test was conducted to analyze the pre-posttest scores of the three groups to determine whether the students’ performance was improved with the intervention of the computerized dynamic assessment system.

**Degree of use and completion**

To understand the treatment intervention and whether differing degrees of dynamic assessment use significantly affected learning outcomes, the degree of use and completion (high, medium, and low) were employed as independent variables. The posttest scores were used as dependent variables and one-way analysis of variance (ANOVA) was conducted.

**Research findings**

According to the research process, data were collected after the students completed a pretest, three weeks of computerized dynamic assessment treatments, and posttest achievements. Using the information collected, data analysis was conducted considering two dimensions, that is, learning effectiveness when employing the computerized dynamic assessment system, and learning outcomes in relation to the level of computerized dynamic assessment use.

**Learning effectiveness**

In this study, according to their grades for the accounting course from the previous semester, the students were divided into high-, medium-, and low-scoring groups. The intervention (a computerized dynamic assessment system) was used as the independent variable, and the students’ posttest scores were used as dependent variables for conducting a paired t-test. Both the pretest and posttest comprised 16 question items appropriate for the scope of the lesson plan. The total possible score was 16 points. Table 1 shows the paired t-test results of the three groups.

For the high-scoring group, the posttest score (M = 13.82) was significantly higher than the pretest score (M = 9.82; p = .000). For the medium-scoring group, the posttest score (M = 10.55) was significantly higher than the pretest score (M = 6.64; p = .000). For the low-scoring group, the posttest score (M = 7.50) was significantly higher than the pretest score (M = .45; p = .000). These results indicate that for all three groups, the computerized dynamic assessment system significantly benefited the students’ learning performance for the Accounting Entry course. Using
the pretest – train – retest – train – retest…posttest model and expert concept map, the students were able to understand the concept of Accounting Entry and cultivate problem-solving skills in all three levels.

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>11</td>
<td>9.82</td>
<td>2.04</td>
<td>13.82</td>
<td>1.25</td>
<td>8.563**</td>
</tr>
<tr>
<td>Medium</td>
<td>11</td>
<td>6.64</td>
<td>2.69</td>
<td>10.55</td>
<td>2.54</td>
<td>5.253**</td>
</tr>
<tr>
<td>Low</td>
<td>12</td>
<td>4.50</td>
<td>2.68</td>
<td>7.5</td>
<td>3.50</td>
<td>3.555**</td>
</tr>
</tbody>
</table>

** p < .01 , *** p < .001

**Learning profile and learning styles**

As mentioned previously, one extremely important reason for developing a computerized dynamic assessment system is to identify students’ learning processes and provide appropriate guidance. Thus, by analyzing student learning profiles, student learning styles can be identified and appropriate follow-up processes provided accordingly.

*The high-posttest students tended to be sequential learners*

The high-posttest scorers (14 questions correct out of 16) were 9 students. The process profile of Student S03 is shown below:

1. Practiced all Accounting Classification units (G1-G5)
2. Attempted the Accounting Entry unit assessment, but did not pass
3. Ignored the system feedback
4. Practiced all Principle of Duality units (G6-G10)
5. Passed all assessments for the Principle of Duality unit
6. Passed all Accounting Entry unit assessments

According to the profiles, 7 students (high posttest scorers group) had a similar learning process pattern. According to Felder and Silverman (1988), these students were most likely sequential learners, who conduct linear thinking and problem solving sequentially. Sequential learners are adept at convergent thinking and analysis. They are considered to fully understand the learning materials, processes, and prepare appropriately. Therefore, sequential learners can effectively solve difficult and complex problems.

*The low-posttest students tended to be global learners*

Felder and Silverman (1988) stated that global learners tend to jump between concepts and feel out of sync with fellow students. This process of learning generally matched the pattern exhibited by the low-posttest scoring group. The low-posttest scorers (8 questions correct out of 16) were 8 students, who all experienced difficulty with the Accounting Classification and Principle of Duality units. During the assessment, all 8 students completely ignored the system feedback. Some attempted to complete the same unit and continued to fail because concepts were unclear. Meanwhile others wasted their time changing and switching between units. However, during the last week of this study, a number of the students were finally willing to follow the feedback and guidance provided by the system and returned to the directed units for re-practice. The results showed that the students who were willing to try achieved slight progress; however, it was too late in the course to have a significant difference.

Based on the students’ use experience and patterns, the dynamic assessment system is effective for students with various learning styles; however, learning performance is only enhanced if the system feedback and guidance are accepted. Moreover, global learners require additional time to familiarize themselves and adjust to the system. “They may feel stupid when they are struggling to master material while most of their contemporaries seem to have little trouble” (Felder & Silverman, 1988, p. 679); however, with sufficient time and patience, they can eventually master the skills. For this case, the learning process requires at least 3 to 4 weeks to achieve a noticeable difference.
Learning outcomes for the use levels

In addition to the learners’ summative scores, this section describes the analysis results of how student learning processes lead to differing learning outcomes. The following two themes were explored: (1) Whether students with differing use degrees achieved different learning outcomes; and (2) whether students with differing completion degrees achieved differing learning outcomes. The degree of use was based on the accumulated number of passed learning units. According to the students’ learning profiles, the highest value was 59 and the lowest was 11. The top 30% were classified into the highest group, the bottom 30% were the lowest group, and the remaining learners were classified into the medium group. The degree of completion was determined by the completion of each major concept area. Learners who completed one concept area (e.g., principle of duality) and passed all units were awarded 1 point. If the learners completed all three concept areas but did not pass all units, they still received 1 point. Based on the total numbers, the top 30% was classified as the highest group, the bottom 30% were the lowest group, and the remaining learners were classified as the medium group.

Learning outcomes for various degrees of use

According to the degree of use, the students were divided into high-, medium-, and low-usage groups, which were used as independent variables. The students’ posttest scores were used as dependent variables, and one-way ANOVA and a Scheffé post hoc comparison were conducted. Table 2 shows the varying degrees of use for each group. Table 3 is a comparison of the posttest scores for varying degrees of use.

Table 2. The description of groups in different degree of using the computerized assessment system

<table>
<thead>
<tr>
<th>Degree of using</th>
<th>N</th>
<th>Mean</th>
<th>Std.</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>11</td>
<td>12.64</td>
<td>1.80</td>
</tr>
<tr>
<td>Medium</td>
<td>10</td>
<td>10.90</td>
<td>3.35</td>
</tr>
<tr>
<td>Low</td>
<td>13</td>
<td>8.46</td>
<td>4.14</td>
</tr>
<tr>
<td>Total</td>
<td>34</td>
<td>10.53</td>
<td>3.67</td>
</tr>
</tbody>
</table>

Table 3. The comparison of posttest scores in different degree of using the computerized assessment system

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Scheffé</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between groups</td>
<td>105.794</td>
<td>2</td>
<td>52.897</td>
<td>4.842*</td>
<td>High &gt; Low</td>
</tr>
<tr>
<td>Within group</td>
<td>338.676</td>
<td>31</td>
<td>10.925</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>444.471</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05

Table 2 shows that the high-use students who employed the dynamic assessment system achieved the highest posttest scores (M = 12.64). The posttest score for the medium-use students was 10.90, and that for the low-use students was 8.46. Furthermore, the ANOVA results showed that the posttest scores for the varying degrees of use differed significantly between groups (F = 4.842; p = .015). The Scheffé post hoc analysis results show that the posttest scores of the high-use students were significantly higher than those of the low-use students. The results indicate that additional practice using the computerized dynamic assessment system can enable students to improve their learning outcomes.

Learning outcome for various degrees of completion

To determine the learning outcomes for varying completion rates between the pretest and posttest, a completion rate calculation method was employed. If the students practiced and passed all the units of one subject area, this was considered one completion. If the students practiced and passed all the units of the three subject areas, this was considered three completions. If the students spent time practicing, but did not pass all three subject areas, this was considered one completion.

According to their degree of completion, the students were divided into high, medium, and low degrees of completion groups, which were used as independent variables. The progress magnitude between the pretest and posttest was used as the dependent variable. Furthermore, one-way ANOVA and Scheffé post hoc comparisons were conducted. Table 4 shows the varying degrees of completion for each group.
Table 4. The description of groups in different degree of completion the computerized assessment system

<table>
<thead>
<tr>
<th>Degree of completion</th>
<th>N</th>
<th>Mean</th>
<th>Std.</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>7</td>
<td>1.57</td>
<td>2.57</td>
</tr>
<tr>
<td>Medium</td>
<td>8</td>
<td>2.75</td>
<td>2.05</td>
</tr>
<tr>
<td>Low</td>
<td>19</td>
<td>4.74</td>
<td>1.79</td>
</tr>
<tr>
<td>Total</td>
<td>34</td>
<td>3.62</td>
<td>2.37</td>
</tr>
</tbody>
</table>

Table 5 is a comparison of the pretest and posttest scores for varying degrees of completion. The results indicate that students with varying degrees of completion differed significantly regarding progress magnitude ($F = 7.223; p = .003$). The Scheffé post hoc analysis results showed that the progress of the high degrees of completion students was significantly greater than that of the low completion students. The results also showed that by using the computerized dynamic assessment system and consistently repeating the practice items, the students could clarify their understanding of the concepts and improve their learning outcomes.

Table 5. The comparison of the progress between pretest and posttest scores in different degree of completion

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Scheffé</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between groups</td>
<td>59.131</td>
<td>2</td>
<td>29.565</td>
<td>7.223**</td>
<td>High &gt; Low</td>
</tr>
<tr>
<td>Within group</td>
<td>126.898</td>
<td>31</td>
<td>4.093</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>186.029</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** $p < .01$

Analysis of learning difficulty for various units

In addition, to investigate the learning progress flow and error patterns for Accounting Entry after employing the dynamic assessment system, learning difficulty analysis of two dimensions was conducted: (1) Employing the degree of use to determine students’ learning difficulties; and (2) employing the pretest and posttest error types to examine learning problems.

Employing the degree of use to determine students’ learning difficulties

The results of analyzing student learning profiles show that the students tended to practice and complete the units they felt were easy while avoiding difficult units. The five units employed were assets (193 times), revenue (189 times), liabilities (161 times), loss (155 times), and owner equity (153 times) (see Table 6). The list of learning profiles matched the general teaching experience of accounting instructors. Liabilities, loss, and owner’s equity were the more difficult units; thus, the students tended to avoid them and achieve lower scores. Guidance and positive feedback from the instructors was critical to encourage the students to continue learning.

Table 6. The usage sorting list of all five units

<table>
<thead>
<tr>
<th>Unit</th>
<th>Assets</th>
<th>Revenue</th>
<th>Liabilities</th>
<th>Loss</th>
<th>Owner’s equity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usage (passed number)</td>
<td>193</td>
<td>189</td>
<td>161</td>
<td>155</td>
<td>153</td>
</tr>
</tbody>
</table>

Using the learning progress flow pretest and posttest error patterns to explore learning problems

Analysis of the pretest results (before the intervention of the dynamic assessment system) show that the number of students who selected to correct answer for Items 1, 3, 7, 8, 13, 15, and 16 was less than 10 (less than 1/3 of all the students), and the item type included owner equity, assets, liabilities, and loss. Items 8, 13, and 15 were regarding complex journalizing, which is a relatively difficult subject for students.

After the intervention of the dynamic assessment system, the students’ posttest scores were significantly higher than those of the pretest. According to the posttest results, the correctly answered items, which was achieved by 17 students (less than half of the total students), were Items 7, 8, and 13. The item types included loss, and reduced
assets and liabilities. Item 13 covered complex journalizing. The posttest results identified the items that were
difficult to learn and indicated that the rate of correct answers was greatly enhanced following the intervention. Items
analysis provided both the students and instructors with further understanding and preparation.

Conclusion and recommendations

In this section, we consolidate the findings of this study to form our conclusion. In addition, based on the student
learning profiles and teaching diagnosis, specific recommendations are proposed as a reference for future
implementations of dynamic assessments in vocational high school curriculum design.

Conclusion

According to the assessment results from the previous semester, the students were divided into high, medium, and
low groups. After the intervention of computerized dynamic assessment system, the posttest scores were
significantly higher than the pretest scores for all three groups. This indicates that the computerized dynamic
assessment system can improve the learning outcomes for students of all levels.

Analysis of student learning profiles showed that the learning style of high-posttest scorers was most likely
sequential learners, who develop linear thinking and problem-solving skills gradually, which leads to better learning
results for this particular design. The low-posttest scores tended to be global learners with unclear learning goals,
who also tended to ignore the system feedback. They continually changed and switched between learning units
without fully understanding the basic knowledge and wasting time with unnecessary information. However, the
analysis results also indicated that if the students were willing to follow the feedback and guidance and allocated
enough time, they could achieve good progress. This matched the learning and teaching style for global learners
(Felder & Silverman, 1988).

Based on the degree of use of the computerized dynamic assessment system, the students were divided into high,
medium, and low groups. The analysis results of the posttest scores show significant differences between the three
groups. The posttest scores for the high degree of use group were significantly higher than those for the other two
groups. The results indicated that high-intensity learners performed better, and the more the students used the system,
the more it enhanced learning.

Comparing the degrees of completion, students with high degrees of completion performed significantly better than
those with low degrees of completion regarding the posttest-pretest progress. The results confirmed that the
computerized dynamic assessment system can improve students’ learning outcomes and the higher the degree of
completion, the better the subsequent progress.

Recommendations

The findings of this study showed that the computerized dynamic assessment system offers obvious benefits for
vocational high school students learning accounting using traditional paper and pencil-based assessments. However,
we provide a few suggestions regarding implementation in classroom settings and integration in instructional design.

According to the results of this study, a computerized dynamic assessment system benefits students of all levels. The
higher the degree of completion, the higher the learning outcomes. Instructors are recommended to integrate the
system in teaching. A training session for both the teachers and students is also recommended for successful
implementation. The dynamic system helps to narrow the achievement gap of the students.

The main structure of the computerized assessment system is the continuum of assessment model, which includes
evaluation functions, and can identify teaching and learning problems. During the assessment process, according to
the types of errors, the system can identify unclear concepts or inadequate problem-solving skills and provide
feedback and guidance. Instructors can use this system to realize the learning conditions of students and provide
appropriate remedial instruction. Instructors should emphasize the development of students’ learning potential.
Using the expert concept map as the framework, the computerized dynamic assessment system provides an adaptive learning capabilities system. Therefore, it can be used for individual learning units and summative assessments. Using this system, instructors can obtain achievement scores, and more importantly, the students' learning processes and potential can be identified by instructors. Thus, this system can provide critical information for teaching and learning.

Acknowledgements

This study is supported in part by the National Science Council of the Republic of China, under contract numbers NSC 98-2511-S-024-006-MY2 and NSC 100-2628-S-024-002-MY3.

Reference


A Novel Agent-Supported Academic Online Examination System

Tuğrul TAŞCI1*, Zekeriya PARLAK2, Alpaslan KİBAR3, Nevzat TAŞBAŞI1 and H. İbrahim CEBECİ1

1Faculty of Informatics of Computer Engineering Department, Sakarya University, Turkey // 2Faculty of Engineering of Mechanical Engineering Department, Sakarya University, Turkey // 3Distance Education Center, Sakarya University, Turkey // ttasci@sakarya.edu.tr // zparlak@sakarya.edu.tr // kibar@sakarya.edu.tr // ntasbasi@sakarya.edu.tr // hcebeci@sakarya.edu.tr

*Corresponding author

(Submitted August 9, 2012; Revised December 8, 2012; Accepted March 13, 2013)

ABSTRACT

Proper execution of exams aimed at assessment & evaluation is of critical importance in Learning Management Systems (LMS). Problems arising from human-centered errors or technical difficulties may lead to questioning of exams, and thus of reliability and efficiency of the distance education systems. Online examination system architecture proposed in this paper provides for integrated management of main functions such as question pool creation and update, exam authoring, execution and evaluation, and management of the feedbacks from students, along with ensuring use of analysis reports related to the questions and exams created by an intelligent agent in the decision-making processes. After conducting analyses on the distance education (DE) system of Sakarya University, it was found that the proposed intelligent agent supported online exam system detects the problems that arise to a large extent and enables the instructors to decide more easily in a shorter time. This system, which is possible to expand with its flexible structure to include additional intelligent features with the aim of resolving different problems, can also be adapted to different institutions that use online examination systems.

Keywords

Architectures for educational technology system, Distance education and telelearning, Evaluation methodologies, Intelligent tutoring systems

Introduction

Following the rapid developments in the information technologies, online education gains importance as an alternative to the traditional teaching models. In online education, teaching independent of time and space constraints brings a fast increase in the numbers of students. For evaluation of such high numbers of students, online test methods are more used than classical methods (Zhang et al., 2006). Structural problems of classical evaluation such as requiring too much time (Gawali & Meshram, 2009) and not allowing sufficient means for student and instructor feedbacks (Guo & Mao, 2010; Zhang et al., 2006) can be eliminated owing to the functions of online test system such as efficient management of evaluation process, feedback system and diversity and pace of student evaluation (Tallent-Runnels et al., 2006). Studies on online examination systems constitute a major part of the studies on online education in the recent years (Tallent-Runnels et al., 2006). In many of these studies, theoretical models relating to the exam systems are provided along with application approaches. In their study, Mustakerov & Borissova (2011) predicted that the two-layered examination system they offered ensured significant benefits owing to the flexible adjustment approach during the testing process. Authors also asserted that the system they prepared could be used for formal and informal evaluation processes. And in other theoretical studies in the literature, examination systems are designed with 3 or more layers with the use of broader architectures. Although featured with differing names in the literature, the examination system is composed of preliminary preparation layer that covers administrator appointments and student enrollments (Brusilovsky & Miller, 1999; Gawali & Meshram, 2009; Keleş et al., 2009; Aye & Thwin, 2008; Guo et al., 2008; Zhang et al., 2006), exam preparation layer consisting of questions, question pools and exams (Brusilovsky & Miller, 1999; Gawali & Meshram, 2009; Keleş et al., 2009; Jun, 2009; Aye & Thwin, 2008; Guo et al., 2008; Li & Wu, 2007; Zhang et al., 2006) and evaluation layer where the results of the exams are calculated (Brusilovsky & Miller, 1999; Gawali & Meshram, 2009; Keleş et al., 2009; Jun, 2009; Li & Wu, 2007; Zhang et al., 2006). Aimin & Jipeng (2009) designed the three layers as administrator, student and instructor, whereas Jin & Ma (2008) created the exam architecture by the help of student and instructor layers alone.
Appropriate management of feedbacks received from students and reporting and analysis processes performed on the exam results are important for proper execution of the examination systems. In his study on collaborative learning process, Wilson (2004) investigated the extent the feedback process affects the student performance by comparing with the students who receive classical face-to-face education. The author emphasized that student feedbacks may render evaluation process more efficient. Crisp & Ward (2008), on the other hand, put that getting feedbacks in the systems with too many students can only be achieved by help of online systems. In their study, Shen et al. (2001) emphasized the importance of adjustment and update processes based on student feedbacks for proper execution of online exams. Also, the authors designed an intelligent monitoring system with automatic reporting and filtering capability. In another study, where importance of the monitoring systems is emphasized (Tinoco et al., 1997), a module, in which the exam results are evaluated, reported and comments derived from these reports by help of analyses, is added to the examination system architecture. Analyses on exam result are also included. Hang (2011) expressed that an online examination system architecture must include re-examination, automatic scoring, question pool update, flexible question design and system security topics.

Architectural flexibility and strong information technology infrastructure of online examinations systems make it likely for them to possess processes suitable for intelligent approaches. In this context, there is high number of studies in the literature based on intelligent approaches or supported by them. Keleş et al. (2009) added two intelligent agents named Advisor and Planner to their learning platform called ZOSMAT, which they designed to support students online and face-to-face. Advisor agent makes a proposal by mainly analysing the test results and reports and by taking the education science facts into consideration, while the Planner agent matches the information received from the Advisor agent with the system resources. In the decision making process, Advisor agent employs a rule-based expert system approach. In the agent-based study by Jin & Ma (2008), formation of the exam and evaluation of the results are carried out by two different agents. Alexakos et al. (2006) likewise executed evaluation process by the help of intelligent agents and used genetic algorithms along with Bayesian Networks for these processes. Unlike other studies, Gawali & Meshram (2009) designed the examination system in its entirety, but no in part, on the basis of three agents, which are Stationary, Mobile and Main. Aimin & Jipeng (2009) on the other hand incorporated intelligent approaches to the exam system architecture through linear algorithms that make random choices in the question selection and exam formation processes, without making use of agent technology.

Proper evaluations of students affect their learning performances (Wang et al., 2008). It can be said that elimination of human and/or system centered errors in the examination systems and good execution indirectly affect student performance and satisfaction. In this regard, it is important to design an auxiliary system or an intermediary layer that will identify the possible difficulties in the main examination processes via reports and student feedbacks and generate solution proposals. As a result of effective use of information technologies during this process, it can be said that intelligent agents can be used for reporting and filtering works in the intermediate layer (Shen et al., 2001). Also, it appears as a necessity that the system gives the student another exam entry right in case of potential problems, update of erroneous or deficient questions in the question pool, ability to update the scores in case of a correction and such components as flexible question design (Hang, 2011). When all these literature findings are considered, there is an obvious need for design and implementation of an integrated and flexible examination system architecture possessing an intelligent intermediate layer, which is capable of producing solutions for differing problems.

In this study, an integrated online exam system architecture that includes Administration, Implementation and Finalization main layers and a Support Layer (SL) is proposed. Owing to this architecture, integrated management of concurrent changes triggered in the processes on all related layers, depending on the decisions made in the SL (like update of question text and difficulty level; giving the student another exam entry right; and update of scores and reports) are ensured. Also, a Monitoring Agent (MA) is designed to help students through creating reports by analyzing the online exam results in a way that enables students to review all possible problems and make decisions more easily, in order to reduce the problems that arise in the implementation process of the online examination system to a negligible level.

Primary contribution of this study is the proposed online examination system having an integrated and flexible structure. The system manages a number of functions such as collection of feedbacks from students, generating
reports and submitting them to the managers and providing proposals in the decision-making processes by help of analysis within a single module as distinct from the other studies found in the literature.

Agent-supported academic online examination system architecture

Agent-Supported Academic Online Examination System (ASOES), the subject matter of this study, is designed as a basic component of the Learning Management System called AkademikLMS (SAU.PORT) that aims to execute online exams of the distance education programs of Sakarya University. While developing ASOES, the goal was to alleviate the potential problems that could be encountered in the online exams and make the exam management process easier.

ASOES, which is developed to ensure effective management of assessment and evaluation processes of academic institutions that have many distance education programs and to offer ease of use to the students and the instructors, has an architecture comprising three main layers and one support layer as shown in Figure 1. Operating process of ASOES starts with an integration application that automatically replicates course, user and user-course enrollment information from databases into an external system to the ASOES database at certain intervals. The Administration Layer (AL) covers such operations as creation of distance education program, courses and users of this program, as well as definition of course-user associations. Question and exam question pools are created and the exams are published for registered students in the Implementation Layer (IL). Finalization Layer (FL) covers calculation and update of exam results and creation of reports relating to these. Support Layer (SL) possesses a Help Desk (HD) operated on the basis of problem feedbacks and an exam monitoring module that analyzes the system reports and forwards the results to the Help Desk.

![Figure 1. ASOES architecture](image)

**Administration layer**

![Figure 2. Administration layer structure](image)
Within the AL, using the information automatically replicated from the external system to the ASOES database at certain periods, distance education programs are created in the AL, and course and users are registered in these programs. Then, data relating to student-course and instructor-course associations are used for matching. Associations of students with the courses they take are defined by Enrollment, and associations of instructors with the courses they give are defined by Management records (See Figure 2). Each student has to be enrolled for the course related to the exam so that they can take the exam and apply to the HD for problems; and this registration information is sent to Publishing process in the IL and Help Desk process in the SL through the Enrollment record. Likewise, in order for each instructor to create an exam and manage the HD, information of the course relating to the exam is sent to the Authoring process in the IL and Help Desk process in the SL through the Management record.

**Implementation layer**

![Figure 3. Implementation layer structure](image)

As seen in Figure 3, IL is composed of Authoring process where the question and exam question pools are created and their features are defined and the Publishing process where the exams are published.

Users with the Instructor authority coming through the Management record from the AL, add the questions to the question pools of their courses by entering them in the Question Pool Creation sub-process together with their features. Features such as question type, difficulty level, keyword and shuffle for each question are entered by the instructor.

Exam Formation sub-process is another part of the Authoring process. In this process, basic exam features come optionally from the AL for each distance education program. If such information does not come, they are created by the instructor with Custom Features in this process. Later, exam question pool is created and recorded in the Exam Features and Exam Question Pool Database together with the exam features.
In the authoring process, if there is a decision taken to change question text, answer choice or question category as a result of the problems notified by the MA or came to the HD based on student problem feedbacks after performing the exam, these will again be updated by the instructors so that problems in the coming exams are prevented.

In the Publishing process, exam question pool and the defined exam features are combined by the Enrollment record from the AL and the exam is published for each student who start a session.

Support layer

In an online education system, there is possibility that tens of thousands of online exams are performed. Also, it is always possible to face human-centered or technical problems in online exams. When these are taken into consideration, auxiliary tools are needed to overcome the difficulty of administration of the exams and to identify and resolve the problems that arise. SL, which is designed for this purpose, is used for smooth execution of the exams (See Figure 4). SL is composed of MA process creating reports and HD process where these reports and student feedbacks are evaluated.

Monitoring agent

It is seen from literature that e-learning platforms need having a kind of item analysis module in order to execute large numbers of online exam sessions smoothly. Post & Hargis (2012) indicated that providing item analyses and statistics are essential for any examination system. In another study it is emphasized that the information provided by the item analysis assists not only in evaluating performance but in improving item quality as well (Fotaris et al, 2010). Online examination modules of the popular LMS platforms (Blackboard, Canvas and Moodle) are evaluated within this context. Blackboard LMS has built-in item analysis module which produces question and exam reports with discrimination, difficulty, average score, standard deviation and standard error statistics. Moodle has an add-in type item analysis module which is capable of presenting reports similar to Blackboard with additional statistics including skewness and kurtosis of the grade distribution, coefficient of internal consistency and discriminative efficiency. Canvas, has no item analysis module in spite of incoming requests in this direction.
The proposed system with monitoring agent, on the other hand, does not only provide reports and statistics but also generate suggestions for the instructors. By virtue of these suggestions, an instructor gains opportunity to give additional exam-entry right to a particular student, and to change the difficulty level or to update body and options of a specific question. Compared to the other platforms, one of the most significant functions of ASOES is that it creates only the list of problematic questions and exams for the related instructor rather than screening standard reports for every question and exam. Instructors may perform updates in the direction of system suggestions or may insist on their previous decisions by evaluating the related questions and reviewing reports of related students. All of these functions are presented to the instructors within a dashboard-type interface so that they may operate with ease.

MA processes the reports it receives from the Reporting process within FL, subject to the rules created by expert opinion, and conducts Category, Exam Response and False Exam Analyses.

- **Category Analysis**

In category analysis, information regarding difficulty level of the questions coming from the IL and the item difficulty index calculated by the system is compared. MA calculates the item difficulty index \( p \) as follows (Bachman, 2004):

\[
p = \frac{U_p + L_p}{2n}
\]  

(Eq. 1)

Here \( U_p \) and \( L_p \) gives the number of students that answer the question correctly in upper and lower groups, while \( n \) value shows the total number of students in the groups (Bachman, 2004). Kelley (1939) put forth that upper and lower groups should be selected equally as first 27 % and last 27 % according to success ranking.

After the \( p \) value is calculated, the difficulty level interval of this value is determined. In the literature it is emphasized that determination of the difficulty level intervals should be undertaken by the instructor who prepares the exam by taking the purpose of the exam into consideration (Atılgan et al., 2006; Kubiszyn & Borich, 1990). Difficulty level intervals used by MA are provided in Table 1.

<table>
<thead>
<tr>
<th>Difficulty Level</th>
<th>Difficulty Level Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Easy</td>
<td>0.65 – 1.00</td>
</tr>
<tr>
<td>Moderate</td>
<td>0.35 – 0.65</td>
</tr>
<tr>
<td>Hard</td>
<td>0.00 – 0.35</td>
</tr>
</tbody>
</table>

Report of proposal to change the difficulty level prepared by help of Rule 1 is sent to the HD process.

**Rule 1:** “If the difficulty level interval that corresponds to the \( p \) value calculated by the system is different from the difficulty level determined by the instructor, create a report of proposal to change the difficulty level.”

- **Exam Response Analysis**

In the exam response analysis, distribution of the incorrect answer choices are investigated; and for all false choices, ratio of number of checking (CF) for the false choice to the number of checking (CT) for the true choice is calculated as in Equation 2:

\[
R_{ERA} = \max \left\{ \frac{CF_i}{CT} \right\} \quad i: 1 \text{ to } 4
\]  

(Eq. 2)

If the largest of these ratios are larger than a threshold value (T) then Rule 2 is applied. A report is created according to this rule to review the correct answer, which is then sent to the HD.

**Rule 2:** “If \( R_{ERA} > T \), create a false answer report.”
• **False Exam Analysis**

In false exam analysis, for all situations, except when students end their session voluntarily or the system terminates it at the end of the duration of the exam, the error codes in Table 2 are created with the parameters they are linked to.

### Table 2. Error Codes

<table>
<thead>
<tr>
<th>Error Code</th>
<th>Description</th>
<th>Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>Server based error</td>
<td>P1</td>
</tr>
<tr>
<td>E2</td>
<td>Client-server communication error</td>
<td>P2, P3, P4</td>
</tr>
</tbody>
</table>

### Table 3. Error Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Presence of active exam sessions during time when the server does not reply</td>
</tr>
<tr>
<td>P2</td>
<td>Abandoning the session deliberately by student</td>
</tr>
<tr>
<td>P3</td>
<td>Ratio of number of students experiencing problems to all student entered for the same exam</td>
</tr>
<tr>
<td>P4</td>
<td>The completion status of the student regarding his/her other exam sessions</td>
</tr>
</tbody>
</table>

Rule, that includes the parameters in Table 3 and the error codes, is inferred as follows:

\[
\text{Rule 3:} \quad \begin{cases} 
\text{IF } P1=1 \text{ THEN } S\{\text{NEER}\}=\text{Yes} \\
\text{IF } P2=1 \text{ THEN } S\{\text{NEER}\}=\text{No} \\
\text{IF } P3>T1 \text{ THEN } S\{\text{NEER}\}=\text{Yes} \\
\text{IF } P4>T2 \text{ THEN } S\{\text{NEER}\}=\text{Yes} \\
\text{ELSE } S\{\text{NEER}\}=\text{No}
\end{cases}
\]

*S\{NEER\}: new exam entry right proposal

Here $T_1$ and $T_2$ are threshold values determined by P3 and P4 parameters. These values are calculated before the exam based on the frequency tables belonging to past exams. Reports prepared as per the inference in Rule 3 are sent to the HD process.

HD mainly includes decision-making processes relating to the problems encountered during exams by receiving information from problem feedbacks notified by the students and the analysis reports prepared by the MA. This decision-making process is undertaken by the instructors in the role of exam administrators. The instructor may receive special reports from the Reporting process if he/she deems it necessary, in order to receive help to decide.

Students who could not complete the exam session or who claim there are errors in the questions choose one of the categories in Table 4. A problem ticket specific to the student is created together with this category information, which is then sent to the relevant decision-making process within HD.

### Table 4. Problem notification category and the relevant decision-making process

<table>
<thead>
<tr>
<th>Problem notification category</th>
<th>Category ID</th>
<th>Relevant decision-making process</th>
</tr>
</thead>
<tbody>
<tr>
<td>False Question</td>
<td>FQ</td>
<td>False Question Validation</td>
</tr>
<tr>
<td>False Exam</td>
<td>FE</td>
<td>New Exam Entry Right</td>
</tr>
</tbody>
</table>

Problems labeled with the FQ category code in the problem database are sent to the False Question Validation decision-making process. Instructor reviews the related question and decides whether it is false or not. If the question is not false, the ticket is closed by sending an information message to the student. If the question is false, then notification is sent to the Evaluation process in the FL for re-evaluation and to the Authoring process in the IL so that the problem in the question is fixed.

**Help desk**

Problems recorded with the FE code in the problem database is sent to the New Exam Entry Right decision-making process. Also, proposal reports specific to the exam session of the student created at the end of False Exam Analysis in the MA are matched with the problem notifications during the New Exam Entry Right decision-making process.
Problems that cannot be detected by the MA and therefore cannot be matched by the problems in the problem database are sent to the New Exam Entry Right decision-making process as a separate report. In the decision-making process, the instructor may give a new exam entry right to the student or close the relevant without giving such right after evaluating the problem. In both of these cases, a notification mail is sent to the student.

Reports automatically generated by the MA as a result of Exam Response and Category analyses are sent to the Analysis Based Question and Category Validation decision-making process. The instructor considers the data from the Exam Response analysis and decides whether the pre-specified answer choices of the questions are false or not; and decides whether the difficulty levels of the questions are defined correctly or not using the data from the Category analysis. He/she sends the update request regarding the correct choice and/or difficulty level of the question to the Authoring process.

**Finalization layer**

As shown in Figure 5, FL consists of Evaluation process where exam result information is processed and turned into scores and Reporting process where all information relating to the exam is processed and turned into reports. In the Evaluation process, answers of the student who has completed the exam are matched with the answer key and a score is calculated for that student. In the Reporting process, all reports needed for the analyses to be made by MA are automatically generated after the exam. Also, all kinds of reports regarding the exam can be created within this process upon request of the instructor. After the reports are evaluated within SL, if re-evaluation decision is made, scores of the student is updated in the Evaluation process.

**ASOES implementation**

Distance education at Sakarya University was launched in 2001 with IBM Lotus LearningSpace. Online examinations of 400 students were hardly conducted because of the huge technical problems notified via e-mail, phone or discussion forums. It was obtained from exam reports that only 65% of exam sessions were started and 84% of those were completed without any error.

Beginning from the year 2003, the number of students was raised to 2500. Therefore, it became a necessity to find an alternative solution due to the unmanageable structure of LearningSpace and expanding institutional requirements. Thus, a new LMS software called SAULMS was developed in 2005 with useful functions such as calendar-based
online exam execution and exam-entry-right management. It was measured that 73% of exam sessions was started during the period (2005-2010) and 94% of those have been finalized by success.

Actually, it was clearly identified by the year 2008 that a high-end LMS software was needed in conjunction with the increase of conducted e-learning programs as well as the number of students. Throughout the searching period, worldwide LMS platforms including IBM LMS, Blackboard, Moodle and Sakai was evaluated. Because those e-learning systems under evaluation have either no customization support or a quite complicated configuration and maintenance process, anew institutional-requirement-specific software (SAUPORT) development attempt was started. SAUPORT has been employed since 2010 for serving more than 20000 students so far in 23 e-learning programs.

Number of online examination sessions today is over 75000 for each semester and execution of those has been performed more efficient than ever before by means of the latest added functions including improved examination calendar, multimedia supported question and options, question bank, elaborative user or course based reporter and advanced exam-entry right manager. Based on the latest reports, it was extracted that the average shares of entrance and errorless completion of exam sessions between 2010 and 2012 were 82% and 99% respectively.

As seen in Table 5, 76156 exam sessions were defined during 2011-2012 fall semester, students started 62280 of these sessions and 61285 of these sessions were completed successfully. In this case, 995 exams were labeled as false exam session.

<table>
<thead>
<tr>
<th>Degree</th>
<th>Total number of exam sessions</th>
<th>Number of started exam sessions</th>
<th>Number of completed exam sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Associate</td>
<td>45318</td>
<td>33095</td>
<td>32433</td>
</tr>
<tr>
<td>Undergraduate</td>
<td>1538</td>
<td>1129</td>
<td>1111</td>
</tr>
<tr>
<td>Master</td>
<td>4829</td>
<td>4076</td>
<td>4026</td>
</tr>
<tr>
<td>Undergraduate completion</td>
<td>24471</td>
<td>23980</td>
<td>23715</td>
</tr>
</tbody>
</table>

Figure 6 shows number of failed exam sessions and the number of problems written by students to the HD. As can be understood from the graph, there are differences in the number of incomplete sessions and the number of messages received. Most important reason behind this is that the student thinks that he/she has completed the exam successfully or he/she does not write anything to the HD despite facing a problem. As it can be seen in the figure, about 2/3 of incomplete exam sessions are not notified as problems by students. By use of the MA application in the proposed system, it was possible to develop a solution to the problems of the students who do not make any notification to the system along with those who make notification.
Exam implementation

Integrated with the external system (Student Affairs database), ASOES is included in the course-instructor/students database. In order to prepare an exam, firstly there should be enough number of questions in the question pool relating to the course. The instructor makes choice of questions from among the questions he/she has earlier entered and publishes the exam. When the student enters into published exam, a session starts. Exam score of the student, who has completed the exam session without a problem, is calculated by evaluating the session data. Students experiencing a problem can notify to HD of their problems.

Problem notified by the students and the results of the analysis made by the MA are evaluated by the instructor and he/she makes his/her decision regarding new exam entry rights, re-calculation of the scores, and update of question text, answer key and difficulty levels of the questions as deemed necessary after which the exam process is finalized and exam results are publicized.

Monitoring agent implementation

MA has three basic functions as Category, Exam Response and False Exam analysis. MA analysis results may be managed on the system by the instructor through a single screen (See Figure 7).

![Monitoring agent](image)

**Figure 7. Monitoring agent**

In the Category analysis, item difficulty index in calculated based on the answers given to the exam questions, and new difficulty level proposal, difficulty level averages of the same question in the previous exams and the difficulty level pre-specified by the instructor are submitted to the instructor. Under light of this information, the instructor can update the difficulty level through the MA interface.

As an example to the Category analysis provided in this study, data of the Computer for Beginners course in the 2011/2012 fall semester of the distance education programs of Sakarya University was used. According to the category analysis performed, MA proposed to change the difficulty level for 9 out of 40 questions belonging to this course (See Table 6).

Proposals to change difficulty level are made based on whether the item difficulty index values are within limits of (Table 1). It is shown in the instructor decision column of Table 6 in bold that the instructor has reckoned 4 out of 9 proposed cases and decided to change the difficulty level. While the instructor decides to change the difficulty level, he/she can consider how much the item difficulty index value converges to difficulty level limits. In the MA interface, calculated difficulty index values are submitted to the instructor together with the new level information.
As can be seen in Figure 8, the 29th question assigned to the Easy category and the 2nd, 3rd and 18th questions assigned to the Moderate category are quite far away from the limits. In such case, it can be seen that instructor’s decision to change the difficulty level is a reasonable approach. For values near limits, the instructor may choose not to change the difficulty level. It can be understood from Table 6 that the instructor decided not to change the difficulty level for questions 5, 20, 33, 38 and 38, as item difficulty indices are very close to limits.

Table 6. Category analysis result table summary

<table>
<thead>
<tr>
<th>Question No.</th>
<th>Difficulty Level Specified by Instructor</th>
<th>Item Difficulty Index*</th>
<th>Difficulty Level Proposed by MA</th>
<th>Decision of Instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>MODERATE</td>
<td>0.211</td>
<td>HARD</td>
<td>HARD</td>
</tr>
<tr>
<td>3</td>
<td>MODERATE</td>
<td>0.121</td>
<td>HARD</td>
<td>HARD</td>
</tr>
<tr>
<td>5</td>
<td>EASY</td>
<td>0.628</td>
<td>MODERATE</td>
<td>EASY</td>
</tr>
<tr>
<td>18</td>
<td>MODERATE</td>
<td>0.112</td>
<td>HARD</td>
<td>HARD</td>
</tr>
<tr>
<td>20</td>
<td>MODERATE</td>
<td>0.322</td>
<td>HARD</td>
<td>MODERATE</td>
</tr>
<tr>
<td>29</td>
<td>EASY</td>
<td>0.513</td>
<td>MODERATE</td>
<td>MODERATE</td>
</tr>
<tr>
<td>33</td>
<td>MODERATE</td>
<td>0.347</td>
<td>HARD</td>
<td>MODERATE</td>
</tr>
<tr>
<td>37</td>
<td>MODERATE</td>
<td>0.343</td>
<td>HARD</td>
<td>MODERATE</td>
</tr>
<tr>
<td>38</td>
<td>MODERATE</td>
<td>0.342</td>
<td>HARD</td>
<td>MODERATE</td>
</tr>
</tbody>
</table>

* Calculated by Equation 1

Figure 8. (a) Distribution of item difficulty index: Easy

Figure 8. (b) Distribution of item difficulty index: Moderate
As the instructors create questions and add them to the question pool together with their answers, they can make errors in assigning the answer choices. It is almost impossible for these students to identify such problems. **Exam Response** analysis is performed on questions that are used for the first time in the system in order to determine if there are answer choice errors.

In the **Exam Response** analysis, choices selected by the students are compared with the answer key from the IL. At the end of the comparison, number of correct answers as well as distribution of the incorrect answers is determined for each question. Ratio of each incorrect choice to the choice defined as correct by the instructor is calculated. It is inquired if this ratio is larger than the threshold value identified as 0.70 based on experience in the previous years. Values for the questions for which proposal of change was made at the end of the analysis concerning questions used for the first time in all online midterm exams in the 2011/2012 fall semester are provided in Table 7.

**Table 7. Exam Response analysis results table**

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Correct Choice</th>
<th>Convergence Ratio</th>
<th>Decision of Correct Choice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>11</td>
<td>27</td>
</tr>
<tr>
<td>5</td>
<td>47</td>
<td>19</td>
<td>10</td>
</tr>
<tr>
<td>53</td>
<td>30</td>
<td>38</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>10</td>
<td>43</td>
<td>17</td>
</tr>
<tr>
<td>9</td>
<td>68</td>
<td>51</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>53</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>47</td>
<td>33</td>
<td>13</td>
<td>15</td>
</tr>
<tr>
<td>10</td>
<td>16</td>
<td>36</td>
<td>9</td>
</tr>
<tr>
<td>0</td>
<td>51</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>41</td>
<td>37</td>
</tr>
<tr>
<td>31</td>
<td>85</td>
<td>7</td>
<td>79</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>26</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>21</td>
<td>3</td>
</tr>
<tr>
<td>62</td>
<td>19</td>
<td>20</td>
<td>26</td>
</tr>
<tr>
<td>57</td>
<td>33</td>
<td>50</td>
<td>32</td>
</tr>
<tr>
<td>68</td>
<td>41</td>
<td>21</td>
<td>61</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>7</td>
<td>27</td>
</tr>
<tr>
<td>59</td>
<td>2</td>
<td>54</td>
<td>1</td>
</tr>
</tbody>
</table>

**Exam Response** analysis was performed for 148 new questions and MA created reports for 19 questions for possible erroneous answer choice (See Table 7). At the end of the review by the instructor, answer choice error was detected in 10 questions, and information was sent to the *Authoring* process in the IL for the update of the answer key, and to the *Evaluation* process in the FL for re-calculation of the scores.
In the *False Exam* analysis, proposal decisions for the exam sessions, which the students could not complete, are created by help of the inference algorithm in the MA (See Rule 3). Here, a four-step algorithm is executed where P1, P2, P3 and P4 parameter values determined by the system are queried for the decision.

In the first step of the algorithm, P1 parameter that corresponds to the presence of active exam sessions during time when the server does not reply is inquired to create giving new right for such exam sessions. Otherwise, it is passed to the second step of the algorithm to check whether the student abandoned the session deliberately (P2). If the session is abandoned deliberately, then the proposal will be not to give new right. Otherwise, we pass to the third step. In order to make a proposal at this stage, T1 threshold value should be calculated by exam-based evaluation of the exam sessions of previous semesters, and P3 parameter for the active exam.

In this study, T1 value is calculated by analyzing 370 exams in the 2010/2011 spring semester. When the frequency values in Table 8 are investigated, maximum false exam session rate was found as 4.88% in 351 exams, which is 95% of all exams, and this value was taken as T1 threshold value.

<table>
<thead>
<tr>
<th>Cumulative Number of Exams</th>
<th>Cumulative / Total Number of Exams (%)</th>
<th>Maximum Error Ratio (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>141</td>
<td>38.1</td>
<td>0.00</td>
</tr>
<tr>
<td>185</td>
<td>50</td>
<td>0.70</td>
</tr>
<tr>
<td>222</td>
<td>60</td>
<td>0.96</td>
</tr>
<tr>
<td>259</td>
<td>70</td>
<td>1.45</td>
</tr>
<tr>
<td>296</td>
<td>80</td>
<td>2.14</td>
</tr>
<tr>
<td>332</td>
<td>90</td>
<td>3.23</td>
</tr>
<tr>
<td>351</td>
<td>95</td>
<td>4.88</td>
</tr>
<tr>
<td>370</td>
<td>100</td>
<td>14.81</td>
</tr>
</tbody>
</table>

If P3 values are larger than T1 threshold value, MA proposes to give a new right, in the other case it is passed to the fourth step of the algorithm. In the fourth step, P4 value for the student whose problem is being investigated should be calculated, along with T2 threshold values derived from frequencies of all P4 parameter values calculated in all exams in the active semester. In this study, T2 value was calculated by analyzing the exam sessions of 5585 students in the 2011/2012 fall semester.

<table>
<thead>
<tr>
<th>Cumulative Number of Students</th>
<th>Cumulative / Total Number of Students (%)</th>
<th>Minimum Session Completion Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4347</td>
<td>77.8</td>
<td>100</td>
</tr>
<tr>
<td>4469</td>
<td>80</td>
<td>97.67</td>
</tr>
<tr>
<td>4741</td>
<td>85</td>
<td>95.92</td>
</tr>
<tr>
<td>5023</td>
<td>90</td>
<td>92.98</td>
</tr>
<tr>
<td>5307</td>
<td>95</td>
<td>86.84</td>
</tr>
<tr>
<td>5585</td>
<td>100</td>
<td>0.00</td>
</tr>
</tbody>
</table>

When the frequency values in Table 9 are investigated, it was found that 95% of the students successfully completed at least 86.84% of the exam sessions they started, and this value was taken as the threshold value (T2).

**Conclusion**

Online exams are used as an assessment-evaluation tool in the distance education systems that have quite a number of students today. For such systems, good execution of exams aimed at assessment & evaluation is of critical. Problems arising from human-centered errors or technical difficulties may lead to questioning of the exams, and thus (reliability and efficiency of) the distance education systems.

In this study, an intelligent agent-supported integrated online examination system architecture is proposed. This architecture provides for integrated management of main functions such as question pool creation and update, exam authoring, execution and evaluation, and management of the feedbacks from students, along with ensuring use of
analysis reports relating to questions and exams created by an intelligent agent in the decision-making processes. This way, all problems that can arise in the operation process of the exam system will be able to reduce to an acceptable level, owing to the ability to decide more easily and rapidly provided by the system to the instructors.

Student feedbacks are important for efficiency of online examinations systems. However, these feedbacks are not always enough to identify all problems. In addition, it is possible that such cases as failure to determine the question difficulty levels appropriately, erroneous answer keys, and incomplete sessions, for which no problem is notified, remain without a solution. And in some other cases, problem feedback does not provide sufficient information to make decisions such as giving new exam entry right, re-calculation of the exam score and changing question content and the answer key. In this regard, a mechanism that generates reports by help of certain rules was deemed necessary. For this purpose, a mechanism named MA was designed. This intelligent agent supported online examination system, ASOES, has been effectively used since 2011 for operation of 23 distance education programs of Sakarya University, aiming at 7000 students studying for associate’s degree, undergraduate degree, undergraduate completion degree and master’s degree.

In practice, it was experienced that instructors became aware of the problems only when students informed them and that they spent too much time to solve these problems. In such case, it is understood from the results of the analyses conducted that intelligent agent-supported online examination system largely identified the problems and allowed the instructors to decide more easily.

In this study, an intelligent agent structure with three functions was designed to meet the needs of Sakarya University, with the aim of solving the most frequent problems. Yet, it is possible to encounter different error conditions in the examination systems depending on the purpose of use of different institutions. The proposed system can be expanded to include additional intelligent functions aimed at solving these different errors with its flexible structure. Also, the intelligent agent structures used in this study can be expanded using semantics and fuzzy inference methods.

References


A Five-Phase Learning Cycle Approach to Improving the Web-based Problem-Solving Performance of Students

Fan-Ray Kuo¹ and Gwo-Jen Hwang²*

¹Center for Faculty Development and Teaching Resources, Kaohsiung Medical University, Taiwan // ²Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taiwan // revonkuo@gmail.com // gjhwang.academic@gmail.com

*Corresponding author

(Submitted July 26, 2012; Revised March 20, 2013; Accepted May 03, 2013)

ABSTRACT

Internet technology has become an important channel for accessing information and solving problems. However, researchers have indicated that most students lack effective strategies for searching for information on the Internet, and hence their learning performance is not as good as could be expected. To address this issue, in this study, a five-phase learning cycle approach is proposed to conduct web-based problem-solving activities. From the experimental results, it was found that the students’ web-based problem-solving performance was improved after participating in the learning activities. Moreover, based on the analysis of 170 elementary school students’ web information-searching portfolios as well as their feedback to the assessments, questionnaires and interviews, it was found that the students’ problem-solving ability and their web-based problem-solving performance were highly correlated, implying that the five-phase learning cycle approach could be able to improve not only the students’ web-based problem-solving performance, but also their general problem-solving ability.

Keywords

Web-based information-searching, Problem-solving ability, E-portfolios, Interactive learning environments

Introduction

High-order thinking abilities are viewed as an important competence in education nowadays. Among these abilities, problem solving ability has long been an objective of education. In the past decades, diverse definitions or conceptions of problem solving have been proposed by researchers and educators. A well-recognized definition is the process of solving problems via understanding them, then devising, carrying out and evaluating plans to solve them (Gagne, 1985; Polya, 1981). For example, Sternberg (1988) indicated that problem-solving ability should encompass six skills: (1) identifying the nature of the problem; (2) choosing problem-solving steps; (3) choosing problem-solving strategies; (4) choosing appropriate information; (5) allocating proper resources; and (6) monitoring the problem-solving process. ITEA (2000) defined problem solving as the process of clarifying a problem, making a plan, implementing the plan, and evaluating the result in order to solve the problem to meet a human need or want. Some researchers have emphasized that “creativity” is one of the critical keys to solving problems (Hwang et al., 2007; Higgins, 1994). Some have further indicated that problem solving is a complex thinking procedure which should include the processes of critical thinking, creative thinking and reasoning, as shown in Figure 1 (Higgins, 1994; Hwang, Chen, Dung, & Yang, 2007).

![Figure 1. Three high-order thinking abilities involved in problem-solving ability](image)

ISSN 1436-4522 (online) and 1176-3647 (print). © International Forum of Educational Technology & Society (IFETS). The authors and the forum jointly retain the copyright of the articles. Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear the full citation on the first page. Copyrights for components of this work owned by others than IFETS must be honoured. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from the editors at kinshuk@ieee.org.
Previous studies have revealed several factors that might affect students’ problem-solving abilities, such as intelligence quality, learning materials, learning methods, problem-solving instruction strategies, and parents’ socioeconomic background (Oloruntegbe, Ikpe, & Kukuru, 2010). Among these factors, learning methods and problem-solving instruction strategies are both considered as key factors that influence students’ problem-solving abilities; consequently, many studies related to these issues have been conducted in recent years (Hwang & Kuo, 2011; Kuo, Hwang, & Lee, 2012; Tsai & Shen, 2009). In addition, researchers have reported correlations among information-searching skills, cognitive structure and problem-solving abilities (Eisenberg & Berkowitz, 1990). Bilal (2002) has further indicated that the lack of effective information-searching strategies and high-order thinking abilities could influence students’ performance in searching for information on the Internet. However, most previous research concerning students’ information-searching and problem-solving abilities has mainly focused on the difficulties students encounter in their web searching, while failing to provide prescription strategies for the improvement of information-searching and problem-solving abilities, including selecting, abstracting and organizing the searched data (Lee, 2012; Laxman, 2010; Walraven, Brand-Gruwel, & Boshuizen, 2009). Therefore, an important educational issue emerges regarding how to guide individuals to construct their own knowledge via the process of collecting data, and analyzing and integrating information on the Internet autonomously. Hwang and Kuo (2011) further called such learning activities that engage students in comprehending a series of questions related to a specific issue, and seeking, selecting, abstracting, and summarizing information on the Web to answer the questions, "web-based problem solving." They indicated that with proper design of the target issue and the related questions, the students could be benefited by linking the knowledge learned from textbooks to real-world contexts.

Consequently, this study proposes a five-phase learning cycle approach to help students foster the abilities required to deal with web-based problem-solving activities. To test the influence of the proposed learning approach, a single-group pretest-posttest design was employed in the study. Such a single-group research design has been employed by many previous studies in investigating the effects of learning approaches and strategies (Liao, 2007; Sak & Oz, 2010; Sipe & Curlette, 1996; Yang, 2007). In addition, we attempt to investigate the critical factors that might affect students' web-based problem-solving performance via the proposed five-phase learning cycle approach based on individuals' e-portfolios recorded during a series of learning activities. Accordingly, a hybrid analysis combining quantitative and qualitative data, including students’ information-searching portfolios, is applied to reciprocally investigate the factors influencing the students’ web-based problem-solving performance.

**Literature review**

**Web-based problem-solving**

Recently, many educational institutes have begun to encourage educators to conduct diverse web-based learning activities in which students are given more opportunities through knowledge construction to improve their web-based problem-solving performance (Lefrancois, 1997). Accordingly, learning methods and problem-solving instruction strategies are both considered as key factors that influence students’ problem-solving abilities (Hwang, Wu, & Chen, 2012); consequently, many studies related to these issues have been conducted in recent years (Hwang & Kuo, 2011; Kuo, Hwang, & Lee, 2012; Tsai & Shen, 2009).

Researchers have found that student problem-solving ability can be facilitated through the integration of problem-solving strategies and computer technologies (Merrill & Gilbert, 2008; Ferreira & Lacerda Santos, 2009; Chiu et al., 2009). For instance, Lo (2009) proposed a computer-mediated communication (CMC) tool to build online communication environments for problem-based learning (PBL) based on the problem solving theory. The empirical evidence showed that students were able to communicate, discuss, and co-build the knowledge from the collected information. With the online communication, they were able to seek solutions to the problems raised in learning activities. Meanwhile, the students were satisfied with the online PBL course. Ferreira and Lacerda Santos investigated the combination of collaborative learning theory and problem solving theory into web-based learning scenarios. Their research results showed that students could improve their problem-solving competences through well-structured collaborative problem-solving activities guided by the teacher. In addition, Yu, She and Lee (2010) investigated the effects of two factors: the mode of problem-solving instruction (i.e., Web-based versus non-Web-based) and the level of academic achievement (i.e., high achievers versus low achievers) on students’ problem-solving ability and biology achievement. Their research findings showed that web-based problem-solving instruction has the potential to enhance and sustain the learners’ problem-solving skills over an extended period of time.
Several studies have further shown that effective strategies can not only effectively improve students' web-based problem-solving performance, but can also increase their positive attitudes toward problem solving (Brand-Gruwel, Wopereis, & Walraven, 2009; Hwang & Kuo, 2011). To develop effective strategies, it is necessary to find the critical factors that affect students' web-based problem-solving performance. Consequently, it has become an important and challenging issue to investigate the factors related to web-based problem-solving (Huang, Liu, & Chang, 2012; Kuo, Hwang, & Lee, 2012).

E-portfolios

Portfolios refer to a representative collection of a student’s work, which often amounts to documentation as a personal learning record (Evans, Daniel, Mikovich, Metze, & Norman, 2006). Learning portfolios are usually developed to fulfill the needs of assessing the quality of the student’s past or ongoing performance. Learning portfolios provide educators with direct evidence of the quality of students’ work and the information for evaluating their work in progress (Zubizarreta, 2004).

Researchers believe that e-portfolios are one of the most effective approaches to understanding the value of an educational model of assessment that can be used to guide independent learning, self-evaluation and reflective practice (Alexander & Golja, 2007). These processes also provide a clear set of criteria for evaluating actual learning outcomes (Garrett, 2011; Herner-Patnode & Lee, 2009). Previous studies have proved that e-portfolios are able to facilitate teachers’ analysis and assessment of students’ learning performance. For example, Hwang, Tsai, Tsai and Tseng (2008) proposed a web-based searching behavior analyzing system called Meta-Analyzer which is able to help teachers analyze students’ learning behavior when using search engines for problem solving by recording the students’ web-based problem-solving portfolios, including the keywords they use to search for information on the web, the web pages they browse, the time they spend on selecting and browsing the web pages, the web pages they adopt to answer the questions, and their answers to individual questions. Teachers or researchers are able to access those e-portfolios via a teacher interface provided by Meta-Analyzer. Chang (2008) developed a web-based portfolio assessment system for assessing students' self-perceived learning performance. His research findings show that the system is more effective for the overall self-perceived learning performance of poorly motivated students than it is for highly motivated students. Akcil and Arap (2009) found that education faculty students had positive opinions about the usage of e-portfolios for educational purposes which increased their motivation to study. Kabilan and Khan (2012) further indicated that e-portfolios as a monitoring tool could help pre-service English language teachers recognize their learning and identify their strengths and weaknesses.

To sum up, e-portfolios provide educators with insights into students’ learning behaviors, and enable them to further assess the students’ learning performance. In addition, they can be used to compensate for the weaknesses of quantitative analysis approaches by contextualizing learners’ behaviors in the web-based learning environment. Consequently, with the support of e-portfolio analysis, the factors affecting students’ learning achievement shall be inferred more accurately in this study.

Design and implementation of the research

Research questions

According to the purpose of the study described in the introductory section, the research questions addressed are listed below:

(1) Could students’ web-based problem-solving performance be significantly correlated to their problem-solving ability via the proposed learning approach?
(2) What sub-abilities relating to web-based problem-solving ability can be found via the students' e-portfolios?
(3) What critical factors affecting the students’ web-based problem-solving performance can be deduced by a hybrid analysis of quantitative and qualitative data including the students’ information-searching portfolios?
Research method and participants

The research study using a single-group pretest–posttest design was carried out on the influence of the five-phase learning approach on web-based problem solving ability. In addition, the students' web-based problem-solving portfolios were analyzed to investigate critical factors via the Dependent samples t-test method. Such a single-group research method has been adopted by several previous studies (Liao, 2007; Sak & Oz, 2010; Sipe & Curlette, 1996; Yang, 2007). For instance, Yang (2007) adopted a one-group pretest-posttest design to investigate the effect of doing history projects on promoting high school students' critical thinking skills; Sak and Oz (2010) also conducted a series of creative thinking activities with a one-group pretest-posttest design.

A total of 170 elementary school 5th graders ranging from 11-12 years old from three elementary schools in southern Taiwan participated in this study. The samples were chosen owing to taking an identical social studies course with very similar progress, and all having had one year of web information searching experience. In addition, the study examined the normal distribution in the pretest of the web-based problem-solving performance and problem-solving ability of the participants, showing that the skewness and kurtosis of their web-based problem-solving performance were .04 and .01 respectively, while those of their problem-solving ability were -.07 and -.97. Furthermore, in the posttest, the skewness and kurtosis of web-based problem-solving performance were - .93 and .34, and those of problem-solving ability were .23 and .19. Thus, the examined results fit the criteria of the normal distribution stated by Kline (1998) of skewness below 3.0 and kurtosis below 10.

A five-phase learning cycle approach

The experiment design was constructed on the basis of the problem-solving theory and constructivism, as shown in Figure 4. A five-phase learning cycle, adapted from Eisenberg and Berkowitz’s Big Six method (Eisenberg & Berkowitz, 1996), was used to scaffold students in the process of adopting keywords, identifying information and abstracting information for a given social issue via Meta-Analyzer.

The experiment was conducted over a seven-week period, including four stages, namely a pre-test, learning system orientation, five weeks of learning activities, and a post-test. Moreover, a total of six sets of constructive questions concerning different social issues, namely 'credit card slaves,' ‘renewable energy,’ ‘the greenhouse effect,’ ‘the garbage problem,’ ‘water shortages,’ and ‘the falling birthrate problem,’ were designed and embedded in Meta-Analyzer to provide the students with web-based problem-solving practice, as shown in Appendix 1. Among these issues, the first, ‘credit card slaves,’ was used as both the pre- and post-test issue for evaluating their web-based problem-solving performance before and after the learning activity, while the other issues were used as training cases in the learning activities.

Prior to the experiment, all of the students were given a demonstration and practice using the Meta-Analyzer system, as well as instruction in how experts solve problems using problem-solving skills. Afterwards, to enhance the students’ web-based problem-solving perceptions and competence, a five-phase learning cycle based on a given social issue was conducted for eighty minutes each week in the web-based problem-solving context. The five phases of the learning cycle are described in detail as follows:
1. Prior knowledge construction
   Before conducting the web-based problem-solving activities, fundamental knowledge concerning the given social issue is instructed in the introductory session.
2. Keyword adoption
   In this phase, the teacher mainly elaborates what keywords or synonyms could be employed based on the given social issue.
3. Information identification
   The third phase of the learning cycle is to guide students to identify relevant web pages, to look for accurate solutions, as well as to criticize and recognize the advantages and disadvantages of solutions before solving the problem.
4. Information abstraction
In this phase, students are taught to judge what pages need to be conserved and how to abstract useful information to solve the problem via retrieving information from the web pages.

5. Thinking elaboration
The last phase of the learning cycle is to require them to express their ideas based on the given issue freely on their own. This phase also offers an opportunity for peers to reflect on what they have done, which is consistent with both Piaget's theory of cognitive development (Piaget, 1970) and with the theory of social constructivism (Vygotsky, 1978).

![Diagram](image)

**Figure 4.** Experiment design in the study

The five-phase learning cycle lasted for 5 weeks during which the students practiced with 5 social issues, one per week. The social studies teacher only provided instruction and scaffolding for the first two issues. The scaffolding was faded out and removed in the last three issues so that the students could learn to solve the problems on their own. After the learning activities for all of the social issues were completed, the post-test of web-based problem-solving performance and post-questionnaires were conducted within a period of sixty minutes. Interviews with students were arranged on the basis of the analysis results of the collected quantitative data for further investigation of the factors affecting their web-based problem-solving performance in the proposed learning environment.

**Web-based learning environment**

To collect more practical data from students’ learning portfolios, the web-based searching behavior analyzing system, Meta-Analyzer (Hwang et al., 2008), was used to assist in tracing and analyzing the learners’ information-searching behaviors for a series of questions related to some specific issues. It has been used for analyzing the online information-searching behaviors of learners by previous studies (Chiou, Hwang, & Tseng, 2009; Hwang, Chen, Tsai, & Tsai, 2011). For example, Chou (2008) employed Meta-Analyzer to record students’ information-searching behaviors for investigating the relationships among their informal reasoning, decision-making and online information commitments on socio-scientific issues; Tu, Shih and Tsai (2008) employed Meta-Analyzer to analyze eighth graders’ web searching strategies and outcomes. Recently, Tsai, Hsu and Tsai (2012) used Meta-Analyzer to
investigate the effects of different information searching strategies on high school students’ science information searching performance.

The student interface of Meta-Analyzer consists of three operation areas: The question and answer area is located on the left-hand side of the browser, the information-searching area is located on the upper-right side, and the web pages found by the search engine are given on the lower-right side. To answer the questions, the students can input keywords to search for information and then browse the searched web pages that might be relevant to the topic, as shown in Figure 2.

![Figure 2. Student interface of Meta-Analyzer](image)

Research instruments

The assessment of problem-solving ability refers to the test items proposed by Zhan and Wu (2007). The assessment consists of 10 items for evaluating the students’ ability in terms of “awareness of problems,” “identification of the nature of problems,” “recognition of factors related to the problems,” “identification of more information needed to solve the problems” and “determination of solutions.” The perfect score of the assessment is 20. To ensure inter-rater reliability, three senior social studies teachers from three elementary schools were involved in the rating. Moreover, before the formal rating of the experiment, 32 non-experimental high-grade students participated in the test in an attempt to verify the correlation among the raters. Measured via Pearson correlation analysis, the inter-rater reliability of the pre-test has a Cronbach's $\alpha$ value of 0.915, showing high consistency between the ratings of the various teachers.

The assessment of web-based problem-solving performance refers to one of the question sets proposed by Kuo, Hwang and Lee (2012). It contained four sub-questions embedded in the Meta-Analyzer system and was used to evaluate the students’ web-based problem-solving performance before and after the learning activity. The maximum score of each sub-question is 10, and the perfect score of the assessment is 40. To ensure inter-rater reliability, three senior social studies teachers from three elementary schools were invited to perform the rating. Moreover, before the formal rating of the experiment, a non-experimental high-grade class participated in the test in an attempt to verify the correlation among the raters. The inter-rater reliability was measured with a Cronbach's $\alpha$ value of 0.89 via Pearson correlation analysis, showing high consistency between the ratings of the various teachers.
The questionnaire survey is structured as follows: The first part consists of students’ personal and academic data, while the second part consists of three dimensions for assessing the students’ responses regarding the proposed learning approach. The questionnaire consists of three dimensions with sixteen items in total, including 5 items for task-technology fit developed by Goodhue and Thompson (1995), 5 items for use intention developed by Davis (1989), and 6 items for web-based problem-solving performance originating from the measure proposed by Gagne et al. (1985). These three dimensions are measured with a five-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree). For validation of the questionnaire survey, factor analysis was conducted with 30 students who were non-experimental samples prior to the experiment. Principal component analysis was used as the extraction method to extract 22.76%, 20.73% and 15.69% of variance in rotation sums of squared loadings in terms of dimensions of task-technology fit, use intention and web-based problem-solving performance respectively, and the total variance explained reaches 59.18%. After the experiment, the data of a total of 170 fifth graders were analyzed with factor analysis to measure the sampling adequacy, in which KMO provides an index of proportion of variance among variables. The results of the analysis showed that the KMO reached 0.92, indicating that the questionnaire items exhibit common variance. Accordingly, internal consistency reliability was performed individually for the three different dimensions. The Cronbach’s $\alpha$ for these dimension presented at least 0.82, meaning that the measuring tools of the study have high reliability. Moreover, these items were modified and adapted through an iterative personal interview process with two senior social studies teachers and one professor to verify the completeness, wording, and appropriateness of the instrument and to confirm the content validity. The Cronbach's $\alpha$ values of the three dimensions are 0.82, 0.87 and 0.88, respectively.

In order to measure the students' e-portfolios compiled during the learning process, thirteen indicators were embedded in Meta-Analyzer to record the students’ online searching portfolios that were employed in the study. These indicators have been widely used in many web-based learning activities for analyzing students' online learning achievement, and allow teachers to evaluate the students’ learning status and adapt instructional strategies accordingly (Chou, 2008; Hwang & Kuo, 2011; Kuo, Hwang, & Lee, 2012; Kuo, Hwang, Chen, & Chen, 2012). These quantitative indicators can be divided into three dimensions, namely keyword-adopting ability, information-selecting ability and question-answering ability (Hwang et al., 2008).

Figure 3 shows an illustrative example of a student’s searching portfolio recorded by Meta-Analyzer, including the student’s identification, the student’s answers to each question, the teacher’s evaluation of each question, the time duration for each operation, and so on. Via browsing the portfolios, researchers or teachers can analyze the detailed searching behaviors of the students.
Results

Correlation analysis of problem-solving outcomes

This study firstly investigated the correlation between the students’ web-based problem-solving performance and their problem-solving ability via employing the Pearson correlation analysis method. Table 1 shows the analysis result of the pre-tests of both learning outcomes. A medium correlation between two dependent variables \( r = 0.317, \ p < .01 \) was found. Moreover, by analyzing the post-tests of both learning outcomes, a close to large correlation \( r = 0.497, \ p < .01 \) was derived, as shown in Table 2. This result could imply that the more the students’ web-based problem-solving performance is promoted, the better their problem-solving ability is.

In this study, the students’ web-based problem-solving performance was improved after a series of training cases, which could be the reason why their problem-solving ability was improved as well. Thus, it is concluded that engaging students in web-based problem-solving activities could foster both their problem-solving ability and web-based problem-solving performance, which is consistent with previous studies (Argelagos & Pifarre, 2012; Hwang & Kuo, 2011; She et al., 2012).

| Table 1. Pearson correlation analysis for the pre-test of learning outcomes |
|-----------------------------|-----------------------------|-----------------------------|
| Dependent variable          | Web-based problem-solving performance | Problem-solving ability |
| Web-based problem-solving performance | 1                           |                             |
| Problem-solving ability     | 0.317**                     | 1                           |
| Note. **p < .01             |                             |                             |

Table 2. Pearson correlation analysis for the post-test of learning outcomes

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Web-based problem-solving performance</th>
<th>Problem-solving Ability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web-based problem-solving performance</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Problem-solving Ability</td>
<td>0.497**</td>
<td>1</td>
</tr>
<tr>
<td>Note. **p &lt; .01</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Analysis of web-based problem-solving portfolios

The students searching portfolios were further analyzed for investigating their web-based problem-solving performance. Paired-sample \( t \) tests were employed to analyze the pre- and post-test records in terms of the thirteen indicators provided by Meta-Analyzer, as shown in Table 3.

<table>
<thead>
<tr>
<th>Table 3. Paired-samples ( t ) tests for the pre-test and post-test of the information-searching portfolios in Meta-Analyzer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ability</td>
</tr>
<tr>
<td>Keyword adoption</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Information Selection</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Question</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Total time for browsing non-adopted pages with revisits taken into account (Sec.)</td>
</tr>
<tr>
<td>Number of marked but not adopted pages</td>
</tr>
<tr>
<td>Number of different adopted pages</td>
</tr>
<tr>
<td>Total time for browsing the different adopted pages (Sec.)</td>
</tr>
<tr>
<td>Number of adopted pages with revisits taken into account</td>
</tr>
<tr>
<td>Total time for browsing the adopted pages with revisits taken into account (Sec.)</td>
</tr>
<tr>
<td>Number of marked and adopted pages</td>
</tr>
</tbody>
</table>

(Note: \( N = 170 \). *\( p < .05 \), **\( p < .01 \), ***\( p < .001 \), #Sec. = seconds)

In terms of keyword-adopting ability, the post-test for the indicators of “average number of keywords used in a search operation” (\( t = 12.40, p < .001 \)), and “number of search attempts for answering the question” (\( t = 3.40, p < .01 \)) had significant improvement over the pre-test.

In terms of information selecting ability, the post-test for the indicators of “total time for web page selection” (\( t = 2.61, p < .01 \), “number of different browsed and non-adopted pages” (\( t = -3.38, p < .01 \), “total time spent browsing the different non-adopted pages” (\( t = -2.96, p < .01 \), “number of browsed non-adopted pages with revisits taken into account” (\( t = -3.72, p < .001 \), “total time spent browsing non-adopted pages with revisits taken into account” (\( t = -2.43, p < .05 \) and “number of marked but non adopted pages” (\( t = -2.16, p < .05 \) were significantly better than in the pre-test. Moreover, in Figure 5, the indicator “total time for web page selection” shows an apparent decreasing tendency in the learning activities, implying that the students were gradually able to shorten the web page selecting time due to adopting more appropriate keywords for the questions. In addition, both indicators “total time for browsing the different non-adopted pages” and “total time for browsing the different non-adopted pages with revisits” exhibit an apparent increasing tendency during the learning activities. These quantitative results show that the students were willing to spend more time browsing different web pages for the correct answers to the questions, implying that their information selecting ability has been significantly promoted.

In terms of question-answering ability, the post-test for the indicators “number of marked and adopted pages” (\( t = -2.17, p < .05 \), “number of different adopted pages” (\( t = -4.45, p < .001 \) and “total time for browsing the different

![Figure 5. Students' information selecting portfolios for a series of web-based problem-solving activities](image-url)
adopted pages” \((t = -2.05, p < .05)\) had significant improvement over the pre-test. These quantitative results show that the students would spend more time referring to different adopted pages to identify answers to the questions, implying that their question-answering ability was significantly promoted.

On the whole, the information-searching portfolios reveal that the participants’ keyword-adopting ability, information-selecting ability and question-answering ability all made noticeable progress according to the pre- and post-test quantitative results, implying that their web-based problem-solving performance significantly improved via a series of learning activities related to social issues embedded in the proposed learning approach.

**Analysis of questionnaires and interview results**

The questionnaire explored task-technology fit, which considers both the task for which the technology is used and the fit between the task and the technology. The questionnaire was evaluated by employing Confirmatory Factor Analysis (CFA) based on three criteria suggested by Fornell and Larcker (1981), namely: (1) All indicator factor loadings should be significant and exceed 0.5; (2) Composite reliabilities (CR) should exceed 0.7; and (3) Average variance extracted (AVE) by each construct should exceed the variance due to measurement error for the construct (i.e., AVE should exceed 0.5). Accordingly, all the standard factor loading values in the confirmatory factor analysis of the measurement model exceed 0.5 and are significant at \(p = 0.001\). In addition, the CR of the constructs range from 0.84 to 0.88, and the AVE ranges from 0.52 to 0.56. Therefore, all three conditions for convergent validity are met, which indicates good internal consistency (Fornell & Larcker, 1981).

The average scores of the questionnaire items ranged from 3.86 to 4.23, as shown in Table 4, implying that most of the students agreed that the web-based problem-solving approach could guide them to solve problems. The questionnaire also explored use intention, which refers to an individual’s subjective likelihood of performing a specified behavior (Ajzen, 1985). The average statistical results of the questionnaire for this aspect ranged from 4.07 to 4.19, indicating that most of the students wished to learn other subjects via this kind of learning mode in the future.

<table>
<thead>
<tr>
<th>Questionnaire</th>
<th>Items</th>
<th>S.D</th>
<th>Mean</th>
<th>(^aFL)</th>
<th>(^aCR)</th>
<th>(^aAVE)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Task-Technology Fit</strong> (adapted from Goodhue and Thompson (1995))</td>
<td>1. I have been provided with the needed functions for completing the learning tasks.</td>
<td>.607</td>
<td>3.95</td>
<td>0.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. The learning approach can guide me to search for relevant information efficiently based on the social issues.</td>
<td>.372</td>
<td>3.86</td>
<td>0.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. The learning approach can guide me to understand the needed information for the social issues.</td>
<td>.532</td>
<td>4.04</td>
<td>0.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. The learning approach can guide me to deeply understand the meaning of social issues.</td>
<td>.540</td>
<td>4.15</td>
<td>0.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5. The learning approach can guide me to figure out how to think about and solve social problems.</td>
<td>.566</td>
<td>4.23</td>
<td>0.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Use intention</strong> (adapted from Davis (1989))</td>
<td>1. I will keep adopting this kind of learning mode in my social studies classes.</td>
<td>.592</td>
<td>4.07</td>
<td>0.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. I will adopt this kind of learning mode to enhance my independent thinking ability.</td>
<td>.542</td>
<td>4.10</td>
<td>0.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. I would like to adopt this kind of learning mode in the future.</td>
<td>.526</td>
<td>4.12</td>
<td>0.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. I would like to recommend this kind of learning mode to my peers.</td>
<td>.530</td>
<td>4.16</td>
<td>0.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5. Overall, I wish to learn other subjects via this kind of learning mode in the future.</td>
<td>.525</td>
<td>4.19</td>
<td>0.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Web-based problem-solving performance</strong></td>
<td>1. I believe this kind of learning mode could enhance both my keyword-adopting and information identifying abilities.</td>
<td>.551</td>
<td>4.25</td>
<td>0.81</td>
<td>0.88</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>2. I believe this kind of learning mode could promote</td>
<td>.543</td>
<td>4.31</td>
<td>0.75</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

178
I believe this kind of learning mode could facilitate me to solve problems on the Internet. 

4. I believe this kind of learning mode could extend my perceptions in analyzing any problem. 

5. I believe this kind of learning mode could help me learn more about social issues. 

6. I believe this kind of learning mode could enhance my understanding of social issues. 

Note. N = 170. a FL: factor loading; b CR: composite reliability; c AVE: average variance extracted

Web-based problem-solving performance refers to an individual’s subjective likelihood of performing a series of web-based problem-solving activities. The average statistical results of the questionnaire for these items ranged from 4.21 to 4.31, showing that a large number of the students positively agreed with and believed that the proposed learning approach applied to the social studies course could not only enhance their understanding of social issues, but could also promote both their problem-solving ability and their information-searching ability on the web.

To derive more insights into the participants’ responses regarding their web-based problem-solving performance, thirty-six students with different levels of achievement were randomly chosen to take part in semi-structured interviews. The interview transcripts were analyzed based on the grounded theory of quantitative analysis (Glaser & Strauss, 1967) in order to infer the critical factors affecting the learners’ web-based problem-solving performance. The grounded theory, a qualitative research approach proposed by Glaser and Strauss (1967), was employed to analyze the interviewed results in the study. It consists of several key analytic strategies: (1) Coding: the process for categorizing qualitative data and describing the implications and details of these categories (e.g., the students’ responses related to the positive effects of the learning approach on their keyword-adopting ability); (2) Memoing: the process of recording the thoughts and ideas of the researcher as they evolve throughout the study (e.g., the researchers took memos by writing down the key words or key points immediately when the students addressed something related to the impacts of the five phase strategies); and (3) Summarizing and generalizing: the process of pulling all of the details together to help make sense of the data by summarizing the findings and determining the most critical factor or point (e.g., by summarizing all of the key points from the students, the researchers were able to find the most important factors affecting the students’ web-based learning performance).

To investigate the students' perceptions of what potential factors would affect their web-based problem-solving performance, the following interview questions were used to elicit their opinions:

1. Do you think the proposed learning cycle approach can benefit you in improving your web-based problem solving? Why?
2. What abilities do you think are being improved in such a learning context? Why?
3. Do you think it is applicable to collocate a series of social issues in the Meta-Analyzer learning system? Why?
4. Would you like to apply such a proposed learning approach in other subjects in the future? Why?

From the interview, it was found that most of the students owed their web-based problem-solving performance to two major factors, that is, "Keyword-adopting and web information-searching abilities" and "Task-technology collocation."

Among the thirty-six students, twenty (55%) indicated that “keyword-adopting and web information-searching abilities” was the most important factor affecting their web-based learning performance. For example, one student indicated that "After participating in the learning activity, I understand the importance of identifying the key points of a problem before solving it. I need to carefully determine the keywords first in order to find useful information on the web." This finding is consistent with previous research indicating that better web learning performance and outcomes could be achieved by improving students’ keyword or synonym identifying and information retrieval abilities (Kane & Trochim, 2004; Tsai, Tsai, & Hwang, 2011; Saunders, 2008).

In terms of Task-technology collocation, eleven of the students (32%) thought that technology-enhanced learning cannot successfully achieve the teaching goal without effective learning strategies and learning tasks, implying the importance of embedding learning strategies and tasks into technology-enhanced learning environments, which echoes the findings of previous studies (Alexiou & Paraskeva, 2010; Casanova, Moreira, & Costa, 2011; Dreyer &
Nel, 2003). For example, one student indicated that "Using this system (i.e., Meta-Analyzer) to realize and solve a series of learning tasks is a good way to enhance my online problem-solving competence." Another student stated that "Social studies issues are well-designed in Meta-Analyzer, which effectively guides me to solve problems in a step-by-step and easy-to-follow manner."

**Discussion and conclusions**

In this study, a web-based problem-solving approach was proposed and used to investigate students’ learning performance via a hybrid analysis of quantitative and qualitative methods, in which the students’ web information-searching portfolios were analyzed to assess their performance. The statistical results show that the students’ web-based problem-solving performance as well as their problem-solving ability was improved via the proposed learning approach. This improvement could be attributed to the proposed approach based on problem-solving theory whereby the students were required to solve social issues following the five-phase learning cycle via the Internet resource. This also implies that the students’ learning performance is affected by interaction with cognitive and social environmental factors, which confirms that technology alone does not cause learning to occur (Bitter & Pierson, 1999). That is, the participants expressed their positive responses regarding the dimension of task-technology fit, showing that a series of social issues embedded in Meta-Analyzer could effectively guide them to figure out how to think deeply about and solve social problems. Moreover, they also gave positive responses regarding the dimension of use intention, indicating that the benefit of the proposed learning environment could strengthen their intentions so that it may be conducted in other subjects. Thus, this corresponds to the perspective of social cognitive theory, in which human behaviors have reciprocal interaction with cognitive, behavioral and environmental influences (Bandura, 1986). In other words, if individual students could obtain support from peers or teachers when conducting a complex learning task, they may gain better learning effectiveness.

Furthermore, the students’ information-searching portfolios revealed that their keyword-adopting, information-selecting and question-answering abilities were significantly enhanced via the series of web-based problem-solving activities. This result is consistent with previous studies (Coles & Robinson, 1989; Higgins, 1994; Kuo, Hwang, & Lee, 2012) which found that keyword-adopting and information-selecting abilities are critical to one's information searching ability on the web. Thus, according to the quantitative and qualitative analyses including the students’ information-searching portfolios, we could deduce that the process of web-based problem-solving performance should consist of critical thinking, creative thinking, reasoning thinking and information-searching abilities, based on the previous studies and current research findings, as shown in Figure 8. That is, students’ information-searching (IS) ability could be the critical factor affecting their web-based problem-solving (WPS) performance. This inferred result of the study could provide a framework for practitioners or educators in the educational field who are engaged in designing web-based problem-solving learning activities.

![Figure 8. The high-order thinking process of web-based problem-solving performance](image)

Although this study provides insights into what critical factors determine students’ web-based problem-solving performance in the proposed web-based learning approach, further investigations might be needed to confirm and extend the inferred results of the study. Thus, several issues need to be considered for future research. First, the extended studies can involve different groups of subjects who experience different learning approaches under more
controlled conditions in order to make more significant conclusions. Second, further study may be needed to add more latent variables relating to web-based problem-solving performance via structural equation modeling (SEM). A hybrid analysis of SEM and e-portfolios is more likely to accurately explain what determinant factors could influence students’ web-based problem-solving performance. Third, further examination of the effects of individual characteristics, such as genders, learning styles or knowledge levels, in web-based learning performance is required. Finally, the learning approach was validated using sample data gathered from three elementary schools in Taiwan. The fact that the participants come from one country limits the generalizability of the results. Other samples from different countries and cultures should be collected to confirm and refine the findings of this study.

Acknowledgements

This study is supported in part by the National Science Council and the Ministry of Education of the Republic of China under contract numbers of NSC 99-2511-S-011-011-MY3, NSC 101-2511-S-011 -005 -MY3 and U19604.

References


Chou, T, W, (2008). The relationships among vocational senior high school students' informal reasoning and decision-making on socio-scientific issues and their online information commitments. (Unpublished master's thesis). National Taiwan University of Science and Technology, Taipei, Taiwan.


Tsai, P. S., Tsai, C. C., & Hwang, G. J. (2011). The correlates of epistemological beliefs toward Internet environments, online searching strategies, and searching outcomes. *The Internet and Higher Education, 14*(1), 54-63.


### Appendix 1

#### Seven sets of constructive questions for problem-solving ability

<table>
<thead>
<tr>
<th>Set No.</th>
<th>Social issues</th>
<th>1st question</th>
<th>2nd question</th>
<th>3rd question</th>
<th>4th question</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Credit Card Slave</td>
<td>How many credit card slaves are there in Taiwan?</td>
<td>What leads them to become card slaves?</td>
<td>What disadvantages and advantages are there when shopping with a credit card?</td>
<td>If you had a credit card, how would you use it to avoid becoming a card slave?</td>
</tr>
<tr>
<td>2</td>
<td>Renewable energy</td>
<td>What are the three forms of power generation used in Taiwan?</td>
<td>In addition to the previous methods, what other methods are there? Give a short introduction to each.</td>
<td>What are the disadvantages and advantages of nuclear power and thermal power?</td>
<td>If you were the Minister of Energy, what form of power generation would you adopt, and why?</td>
</tr>
<tr>
<td>3</td>
<td>Greenhouse effect</td>
<td>What countries are the top two carbon dioxide emitters in the world?</td>
<td>What are the impacts on the Earth of the emission of lots of carbon dioxide?</td>
<td>What solutions can decrease carbon dioxide emissions in life? What can you do?</td>
<td>If you were the Minister of Environmental Protection, what would you do to lower carbon dioxide emissions?</td>
</tr>
<tr>
<td>4</td>
<td>Garbage problem</td>
<td>What are the impacts on the Earth if lots of rubbish is produced? e.g. water, air, soil, etc.</td>
<td>What are three main methods of waste disposal? How do they work?</td>
<td>What are the differences among “landfill,” “garbage incineration,” and “recycling”?</td>
<td>What waste disposal method would you accept to decrease the garbage problem?</td>
</tr>
<tr>
<td>5</td>
<td>Water shortage</td>
<td>How many liters of water are used on average per day in Taiwan?</td>
<td>The annual rainfall exceeds 2,500mm in Taiwan, but there is still a water shortage, why?</td>
<td>Do you think the construction of reservoirs can solve water shortages in southern Taiwan? What impact would they have?</td>
<td>What specific actions can you take to help conserve water at school, at home or anywhere else?</td>
</tr>
<tr>
<td>6</td>
<td>Falling birthrate problem</td>
<td>Please find out the birthrate in 1979 and 2009 in Taiwan.</td>
<td>Currently, what is contributing to the falling birthrate problem in Taiwan?</td>
<td>What industries can be affected by the low birthrate problem?</td>
<td>If you were the President or Premier, what policy would you advocate to promote the birthrate?</td>
</tr>
</tbody>
</table>
Constructive-oriented TPACK. The regression model also juxtaposes the influence of demographic variables against demographic variables such as gender, age, teaching experience and teaching level impact practicing teachers' assessed through the TPACK for Meaningful Learning Survey. Using regression analysis, it examines if This study describes the constructive-oriented TPACK perceptions of 354 practicing teachers in Singapore as the influences of the TPACK constructs of technological knowledge, pedagogical knowledge, content knowledge, relationship between these factors and teachers' constructive-oriented TPACK perceptions have yet to be examined. 2009). While some studies show that teachers' TPACK perceptions were affected by their age, technological knowledge, and pedagogical knowledge (Koh, Chai, & Tsai, 2010; Lee & Tsai, 2010; Pierson, 2001), the studies have only examined teachers' TPACK perceptions with respect to science education, e-learning facilitation, social influence it. Nevertheless, their perceived knowledge gaps in this area are not well understood as published studies derived through a better understanding of their constructivist-oriented TPACK perceptions and the factors that can contribute to their confidence for constructivist-oriented technology integration. The specific challenges faced by experienced teachers and primary school teachers need to be better understood and considered when designing teacher technology professional development.

Keywords
Constructivism, Educational technology, Teacher education, Technological pedagogical content knowledge

Introduction

UNESCO’s ICT competency standards for teachers (UNESCO, 2008) emphasize that teachers need knowledge to use ICT for supporting constructivist learning which involves knowledge construction and problem-solving activities within authentic contexts (Airasian & Walsh, 1997; Duffy & Cunningham, 1996). This can be understood as a kind of technological pedagogical content knowledge (TPACK), which is a term used by Mishra and Koehler (2006) to describe teachers’ knowledge about information and communications technology (ICT) integration. Empirical studies show that practicing teachers do not fully exploit the affordances of ICT tools for constructivist teaching (Lim & Chai, 2008; Starkey, 2010; Webb & Cox, 2004); indicating that constructivist-oriented TPACK could be an area of particular challenge for them. Teachers’ efficacy perceptions had significant positive influence on their adoption of ICT (Wozney, Venkatesh, & Abrami, 2006). Insights for teacher professional development in ICT can be derived through a better understanding of their constructivist-oriented TPACK perceptions and the factors that can influence it. Nevertheless, their perceived knowledge gaps in this area are not well understood as published studies have only examined teachers’ TPACK perceptions with respect to science education, e-learning facilitation, social studies, and mathematics (e.g., Archambault & Barnett, 2010; Graham et al., 2009; Lee & Tsai, 2010; Schmidt et al., 2009). While some studies show that teachers’ TPACK perceptions were affected by their age, technological knowledge, and pedagogical knowledge (Koh, Chai, & Tsai, 2010; Lee & Tsai, 2010; Pierson, 2001), the relationship between these factors and teachers’ constructivist-oriented TPACK perceptions have yet to be examined.

This study describes the constructivist-oriented TPACK perceptions of 354 practicing teachers in Singapore as assessed through the TPACK for Meaningful Learning Survey. Using regression analysis, it examines if demographic variables such as gender, age, teaching experience and teaching level impact practicing teachers’ constructivist-oriented TPACK. The regression model also juxtaposes the influence of demographic variables against the influences of the TPACK constructs of technological knowledge, pedagogical knowledge, content knowledge,
pedagogical content knowledge, technological pedagogical knowledge, and technological content knowledge. The implications for teacher professional development in constructivist-oriented ICT integration will be discussed.

Theoretical background

Constructivism, supported by ICT

Jean Piaget first suggested the term “constructivism” when he observed that children learned by formulating ways to cope with and master their environment (Driscoll, 2000). Constructivism conveyed the notion that reality is in the mind of the learner (Jonassen, 1991) where learning involved meaning construction as learners perceived and interpreted their experiences. Proponents of the social-cultural dimension of constructivism purport that learning occurred as learners interacted with the people and tools of the environment, and deepened their enculturation with the practices of the community (Brown, Collins, & Duguid, 1989; Lave & Wenger, 1991). Pedagogical approaches such as problem-based learning and inquiry-based learning were developed from these conceptions of constructivism.

ICT-supported constructivist learning was first described in Hannafin, Land and Oliver’s (1999) Open Learning Environments which used ICT tools to support information manipulation, problem visualization, and metacognition as students worked with authentic problem contexts. Educational researchers indicated a good interplay between the usage of ICT and the instructional practice of constructivism (Mikropoulos & Natsis, 2011; Tsai, 2001, 2004). The principles underlying ICT-supported constructivist learning were carefully developed by Jonassen, Howland, Marra, and Crismond (2008). Their dimensions of Meaningful Learning stated that constructivist learning with ICT was firstly active, with students manipulating objects in the learning environment and observing its results. Secondly, it was constructive, with teachers engaging students to reflect and articulate their personal understanding of their observations. Thirdly, it involved students in authentic tasks that were based on real-world problems. Next, students were intentional about their own learning as they set their learning goals and plan their problem-solving processes. Finally, it was described as a social process that involved collaborative problem-solving within a classroom community. It was proposed that ICT tools supported these five dimensions as “engagers and facilitators of thinking” (Jonassen, et al., 2008, p. 7). These five dimensions can therefore be used to understand teachers’ constructivist-oriented TPACK.

Constructivist-oriented TPACK in existing TPACK surveys

Mishra and Koehler’s (2006) seven-construct TPACK framework has been used as a theoretical basis for developing surveys to understand teachers’ TPACK perceptions. It comprises of three basic knowledge sources and four others derived from the interaction among these three basic sources. Their definitions are as follows:
1. Technological Knowledge (TK) – knowledge of technology tools.
2. Pedagogical Knowledge (PK) – knowledge of teaching methods.
3. Content Knowledge (CK) – knowledge of subject matter.
4. Technological Pedagogical Knowledge (TPK) – knowledge of using technology to implement teaching methods.
5. Technological Content Knowledge (TCK) – knowledge of subject matter representation with technology.
6. Pedagogical Content Knowledge (PCK) – knowledge of teaching methods with respect to subject matter content.
7. Technological Pedagogical Content Knowledge (TPACK) - knowledge of using technology to implement constructivist teaching methods for different types of subject matter content.

Some aspects of constructivist-oriented TPACK have been addressed in existing TPACK surveys. In Schmidt et al’s (2009) Survey of Preservice Teachers’ Knowledge of Teaching and Technology, constructivist approaches were incorporated as a general question on teaching pedagogies, i.e., “I know when it is appropriate to use a variety of teaching approaches in a classroom setting (collaborative learning, direct instruction, inquiry learning, problem/project-based learning, etc.)”. Graham et al.’s (2009) TPACK in Science survey addressed the active and constructive dimensions of Meaningful Learning through items related to the use of ICT to visualize abstract concepts, e.g., “Help students use digital technologies that allow them to create and/or manipulate models of scientific phenomenon.” Archambault and Crippen’s (2009) survey of teachers’ TPACK for e-learning facilitation also contained items for these two dimensions, e.g., “My ability to create an online environment which allows students to build new knowledge and skills.”; and also the dimension of collaboration, e.g., “My ability to encourage
online interactivity among students.” These three published TPACK surveys do not address the dimensions of authentic tasks and intentionality. On the other hand, Lee and Tsai’s (2010) survey assessed teachers’ self-efficacy for integrating web-based resources and did not examine specific pedagogies.

Chai, Koh, Tsai, and Tan (2011) first designed constructivist-oriented TPACK items by replacing the PK items in Schmidt et al.’s survey with Pedagogical Knowledge for Meaningful Learning (PKML) items constructed with respect to the five dimensions described by Jonassen et al. (2008). For example, the dimension of constructive learning was operationalized as, “I am able to help my students to reflect on their learning strategies.” Its items for TK, CK, PKML, TPK, and TPACK were validated with exploratory factor analysis when administered on 834 Singapore pre-service teachers. The cross-loadings of the TCK and PCK items were attributed to the common reference to content knowledge. A stem, “Without using technology…” was subsequently incorporated into the PCK items so that the applications of content knowledge within and outside a technological context could be better differentiated. Constructivist-oriented pedagogies were also incorporated into the TPK items. An item for the Constructive dimension was stated as, “I am able to facilitate my students to use technology to construct different forms of knowledge representation.” Following these revisions, the survey was again administered on 214 Singapore pre-service teachers (Chai, Koh, & Tsai, 2011) and 455 practicing teachers (Koh, Chai, & Tsai, 2012). The seven TPACK constructs were further validated with adequate model fit. The validation of these survey items paved the way for statistical modeling to examine the factors affecting teachers’ constructivist-oriented TPACK.

Factors affecting practicing teachers’ constructivist-oriented TPACK

Firstly, the inter-relationships among TPACK constructs could impact or were found to be related to practicing teachers’ perceptions of constructivist-oriented TPACK. Chai, Koh, and Tsai (2010) found strong correlations between the TK, PK, CK and TPACK perceptions of Singapore pre-service teachers’ when a general TPACK survey was administered. Similar patterns were observed with respect to teachers’ constructivist-oriented TPACK because Chai, Koh, and Tsai (2011) found that among TPACK constructs, C-PK and C-TPK had the strongest effects on TPACK. These studies have been conducted with pre-service teachers. Practicing teachers, by virtue of their teaching experience, may manifest different types of perceptions which need to be further studied.

Demographic factors, especially age, teaching experience, and gender, may also be related to practicing teachers’ constructivist-oriented TPACK. Case studies of practicing teachers found some relationship between their age and confidence for implementing student-centered ICT-integrated activities (Pierson, 2001). Experienced teachers were also less confident about their TPACK for integrating web-based learning (Lee & Tsai, 2010). Gender has traditionally influenced teachers’ attitudes towards computer use where male teachers tend to be more confident (Teo, 2008). This corresponded with Koh et al. (2010) who found that male pre-service teachers in Singapore perceived higher levels of TK, CK, and Knowledge of Teaching with Technology. Similar gender patterns may apply to practicing teachers. In Singapore, secondary school teachers are recruited to teach the content specialization of their undergraduate major whereas primary school teachers are generalists. This may result in different TPACK perceptions by teaching level which has yet to be examined.

Research questions

Given the above review, a better understanding of practicing teachers’ constructivist-oriented TPACK perceptions is needed. This will be addressed through two research questions:
1. What are Singapore practicing teachers’ constructivist-oriented TPACK perceptions?
2. How do teacher demographics (age, gender, teaching experience, and teaching level) and TPACK constructs (C-TK, C-PK, CK, C-PCK, TCK, and C-TPK) predict practicing teachers’ constructivist-oriented TPACK (C-TPACK)?
Methodology

Study participants

The study participants were 450 practising teachers in Singapore who were attending an ICT professional development programme organized by a teacher education agency. These teachers were nominated by their respective schools to be trained as mentors for the implementation of ICT course design in their schools. The selection criteria given to schools for course nomination were that the course participants should be strong in content knowledge and teaching, but not necessarily in the use of ICT. The survey was administered through a web-based URL that was sent to these teachers by the course administrators to seek their voluntary participation. A total of 354 teachers responded to the survey, constituting a response rate of 78.67%. The survey respondents were largely female teachers (n = 231, 65.25%). The mean age of the study participants were 34.93 years (SD = 6.61). These teachers were fairly experienced as they had an average of 8.83 years of teaching experience (SD = 6.01). About 54% of the teachers (n = 192) were teaching in primary schools while the rest taught in secondary schools or junior colleges.

The TPACK for meaningful learning survey

The TPACK for Meaningful Learning Survey used for this study was constructed from Chai, Koh, and Tsai’s (2011) survey. Jonassen et al.’s dimensions were incorporated more comprehensively to constitute the survey used in this study. Chai et al.’s (2011) TPACK items were revised with Jonassen’s dimensions to improve the assessment of teachers’ constructivist-oriented TPACK. For example, a Constructivist TPACK (C-TPACK) item for the constructive dimension was, “I can structure activities to help students to construct different representations of content knowledge using appropriate ICT tools (e.g., Webspiration, Mindmeister, Wordle)”. Referencing to Shulman’s (1986) conceptualization of PCK, the Constructivist Pedagogical Content Knowledge (C-PCK) items focused on teachers’ facilitation of students’ thinking by addressing their difficulties with content knowledge, for example, “Without using technology, I can address the common misconceptions my students have for my first teaching subject.” Questions about seven types of ICT tools that support knowledge construction (e.g., online sticky notes, mindmapping tools and online visualization tools) replaced the TK items to constitute Constructivist-TK (C-TK) items for better alignment with constructivist teaching. No changes were made to the CK and TCK items because these did not contain pedagogical elements. The TPK items were named more accurately as C-TPK items. The final survey comprised of 32 items. The study participants rated each item on a seven-point Likert-type scale where 1 - Strongly Disagree, 2 - Disagree, 3 - Slightly Disagree, 4 – Neither agree nor disagree, 5 - Slightly Agree, 6 - Agree, 7 - Strongly Agree.

The internal reliability of the survey instrument was high, with an overall Cronbach alpha of 0.96. Internal reliability of the seven TPACK constructs was also established as each had high Cronbach alphas that were larger than 0.90: C-TK (α = 0.94), C-PK (α = 0.94), CK (α = 0.95), C-PCK (α = 0.93), C-TPK (α = 0.95), TCK (α = 0.92), and C-TPACK (α = 0.96). As there were at least ten participants per survey item (Hair, Black, Babin, Anderson, & Tatham, 2010), the number of study participants was considered adequate for exploratory factor analysis which explained a total variance of 83.13%. The factors loaded according to the pre-defined structure for all the seven TPACK constructs with loadings that were larger than 0.50 as recommended by Fish and Dane (2000) (See Table 1). This factor structure was further validated through confirmatory factor analysis using AMOS 19 with satisfactory model fit ($\chi^2 = 1,139.60$, $\chi^2$/df = 2.58, $p < .0001$, TLI = .94, CFI = .95, RMSEA= 0.067, SRMR = .036). The exploratory and confirmatory analyses for this study could not be analyzed with the split-sample approach as there were insufficient survey respondents. Nevertheless, the factor structure of this TPACK survey has previously been progressively validated with Singapore pre-service and in-service teachers (Koh et al., 2012). These results further confirm the validity of the factor structure obtained.

Data analysis

The first research question was examined by analyzing the descriptive statistics of teachers’ TPACK ratings. The second research question was answered by first examining if there were significant Pearson’s correlations between the TPACK constructs, age, and teaching experience. The relationship between demographic variables (gender and
teaching level) and TPACK constructs were examined through independent sample t-tests. The categorical variables of gender and teaching level were then coded as dummy variables. Following this, a stepwise regression model was set-up with C-TPACK as the dependent variable and the remaining TPACK constructs and demographic factors as independent variables.

Table 1. Factor loadings from exploratory factor analysis

<table>
<thead>
<tr>
<th>Factor</th>
<th>Description</th>
<th>Factor loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>TK1</td>
<td>I am able to create web pages.</td>
<td>.66</td>
</tr>
<tr>
<td>TK2</td>
<td>I am able to use social media (e.g., Blog, Wiki, Facebook).</td>
<td>.72</td>
</tr>
<tr>
<td>TK3</td>
<td>I am able to use collaboration tools (e.g., Google Sites, CoveritLive).</td>
<td>.84</td>
</tr>
<tr>
<td>TK4</td>
<td>I am able to use communication tools (e.g., VoiceThread, Podcast).</td>
<td>.83</td>
</tr>
<tr>
<td>TK5</td>
<td>I am able to use online sticky notes (e.g., Diigo, Wallwisher).</td>
<td>.87</td>
</tr>
<tr>
<td>TK6</td>
<td>I am able to use mind tools (e.g., Webspiration, Mindmeister).</td>
<td>.86</td>
</tr>
<tr>
<td>TK7</td>
<td>I am able to use visualization tools (e.g., Wordle, Quizlet).</td>
<td>.80</td>
</tr>
<tr>
<td>C-PK1</td>
<td>I am able to stretch my students’ thinking by creating challenging tasks for them.</td>
<td>.77</td>
</tr>
<tr>
<td>C-PK2</td>
<td>I am able to guide my students to adopt appropriate learning strategies.</td>
<td>.80</td>
</tr>
<tr>
<td>C-PK3</td>
<td>I am able to help my students to monitor their own learning.</td>
<td>.80</td>
</tr>
<tr>
<td>C-PK4</td>
<td>I am able to help my students to reflect on their learning strategies.</td>
<td>.83</td>
</tr>
<tr>
<td>C-PK5</td>
<td>I am able to plan group activities for my students.</td>
<td>.82</td>
</tr>
<tr>
<td>C-PK6</td>
<td>I am able to guide my students to discuss effectively during group work.</td>
<td>.82</td>
</tr>
<tr>
<td>CK1</td>
<td>I have sufficient knowledge about my first teaching subject (CS1)</td>
<td>.77</td>
</tr>
<tr>
<td>CK2</td>
<td>I can think about the content of my first teaching subject (CS1) like a subject matter expert.</td>
<td>.84</td>
</tr>
<tr>
<td>CK3</td>
<td>I am able to develop deeper understanding about the content of my first teaching subject (CS1).</td>
<td>.80</td>
</tr>
<tr>
<td>C-PCK1</td>
<td>Without using technology, I can address the common misconceptions my students have for my first teaching subject (CS1).</td>
<td>.89</td>
</tr>
<tr>
<td>C-PCK2</td>
<td>Without using technology, I know how to select effective teaching approaches to guide student thinking and learning in my first teaching subject (CS1).</td>
<td>.93</td>
</tr>
<tr>
<td>C-PCK3</td>
<td>Without using technology, I can help my students to understand the content knowledge of my first teaching subject (CS1) through various ways.</td>
<td>.91</td>
</tr>
<tr>
<td>C-TPK1</td>
<td>I am able to use technology to introduce my students to real world scenarios.</td>
<td>.64</td>
</tr>
<tr>
<td>C-TPK2</td>
<td>I am able to facilitate my students to use technology to find more information on their own.</td>
<td>.68</td>
</tr>
<tr>
<td>C-TPK3</td>
<td>I am able to facilitate my students to use technology to plan and monitor their own learning.</td>
<td>.74</td>
</tr>
<tr>
<td>C-TPK4</td>
<td>I am able to facilitate my students to use technology to construct different forms of knowledge representation.</td>
<td>.70</td>
</tr>
<tr>
<td>C-TPK5</td>
<td>I am able to facilitate my students to collaborate with each other using technology.</td>
<td>.63</td>
</tr>
<tr>
<td>TCK1</td>
<td>I can use the software that are created specifically for my first teaching subject (CS1). (E.g., e-dictionary/corpus for language; Geometric sketchpad for Maths; Data loggers for Science)</td>
<td>.74</td>
</tr>
<tr>
<td>TCK2</td>
<td>I know about the technologies that I have to use for the research of content of first teaching subject (CS1)</td>
<td>.65</td>
</tr>
<tr>
<td>TCK3</td>
<td>I can use appropriate technologies (e.g., multimedia resources, simulation) to represent the content of my first teaching subject (CS1).</td>
<td>.61</td>
</tr>
<tr>
<td>C-TPACK1</td>
<td>I can formulate in-depth discussion topics about the content knowledge and facilitate students’ online collaboration with appropriate tools. (E.g., Google Sites, CoveritLive)</td>
<td>.65</td>
</tr>
<tr>
<td>C-TPACK2</td>
<td>I can design authentic problems about the content knowledge and represent them through computers to engage my students.</td>
<td>.73</td>
</tr>
<tr>
<td>C-TPACK3</td>
<td>I can structure activities to help students to construct different representations of content knowledge using appropriate ICT tools (e.g., Webspiration, Mindmeister, Wordle).</td>
<td>.73</td>
</tr>
<tr>
<td>C-TPACK4</td>
<td>I can create self-directed learning activities of the content knowledge with appropriate technology.</td>
<td>.73</td>
</tr>
</tbody>
</table>
ICT tools (e.g., Blog, Webquest). C-TPACK5 – I can design inquiry activities to guide students to make sense of the content knowledge with appropriate ICT tools (e.g., simulations, web-based materials).

Findings

Research question 1: What are Singapore practicing teachers’ constructivist-oriented TPACK perceptions?

The teachers studied rated themselves as being above average for all the TPACK categories. They were highly confident of their CK (M = 5.84, SD = 0.93) and their ability to facilitate constructivist instruction, that is, C-PK (M = 5.56, SD = 0.77). Correspondingly, their C-PCK was also high (M = 5.43, SD = 1.05). However, their ratings for constructs involving technology were above mid points, but only close to a rating of five out of the seven-point scale: C-TK (M = 5.17, SD = 0.98), TCK (M = 5.20, SD = 1.09), and C-TPK (M = 5.17, SD = 0.98). Their rating for C-TPACK was less than five points and the lowest among the TPACK constructs (M = 4.86, SD = 1.13).

Research question 2: How do teacher demographics (age, gender, teaching experience, and teaching level) and TPACK constructs (C-TK, C-PK, CK, C-PCK, TCK, and C-TPK) predict practicing teachers’ constructivist-oriented TPACK (C-TPACK)?

Inferential statistics

Table 2 shows that age and teaching experience were not strongly related to TPACK constructs. There was significant small negative correlation between teaching experience and the constructs involving technological knowledge (C-TK, C-TPK, TCK, and C-TPACK). Age was similarly related to these constructs except for TCK where no significant correlation was reported. Age and teaching experience had small positive correlations with C-PCK but were not related to C-PK. The correlations among TPACK constructs had stronger significant positive correlation. Among these, C-TK, C-TPK and TCK had large positive correlations with C-TPACK that were above 0.60.

Table 2. Correlation table

<table>
<thead>
<tr>
<th></th>
<th>Age</th>
<th>Teaching experience</th>
<th>CK</th>
<th>C-PK</th>
<th>C-PCK</th>
<th>C-TK</th>
<th>C-TPK</th>
<th>TCK</th>
<th>C-TPACK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>1</td>
<td>.78</td>
<td>.10</td>
<td>.10</td>
<td>.11</td>
<td>-.26</td>
<td>-.13</td>
<td>-.091</td>
<td>-.21**</td>
</tr>
<tr>
<td>Teaching experience</td>
<td>1</td>
<td>.15**</td>
<td>.09</td>
<td>.11</td>
<td>-.22**</td>
<td>-.16</td>
<td>-.11**</td>
<td>-.25**</td>
<td></td>
</tr>
<tr>
<td>CK</td>
<td>1</td>
<td>.64**</td>
<td>.45**</td>
<td>.33**</td>
<td>.36</td>
<td>.53</td>
<td>.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-PK</td>
<td>1</td>
<td>.31**</td>
<td>.37**</td>
<td>.62**</td>
<td>.51**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-PCK</td>
<td>1</td>
<td>.12**</td>
<td>.15**</td>
<td>.27**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-TK</td>
<td>1</td>
<td>.69**</td>
<td>.68**</td>
<td>.74**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-TPK</td>
<td>1</td>
<td>.67**</td>
<td></td>
<td>.80**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCK</td>
<td>1</td>
<td>.71**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-TPACK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Note. n = 354. ** p < 0.01, * p < 0.05

Independent sample t-tests found significant gender differences only for the constructs that were related to technology, i.e., C-TK, TCK, and C-TPACK (See Table 3). Male teachers rated themselves higher than female teachers for these constructs but the effect sizes were small. Differences by teaching level were only significant for C-TPACK where the primary teachers scored lower ratings than the Secondary and junior college teachers. The effect size was also small.

Table 3. TPACK ratings by gender and teaching level

<table>
<thead>
<tr>
<th>Construct</th>
<th>Gender (n: Male = 123, Female = 231)</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>p (two-tailed)</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-TK</td>
<td>Male: 5.41</td>
<td>1.11</td>
<td>1.11</td>
<td>t(352) = 3.06</td>
<td>0.020*</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>Female: 5.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

190
As aforementioned, stepwise regression analyses were conducted by using C-TPACK as the dependent variable and the remaining TPACK constructs and demographic factors as independent variables. Stepwise regression of the models was statistically significant with an adjusted $R^2$ of 0.75. From Table 4, it can be seen that the technology-related TPACK constructs of C-TK, C-TPK, and TCK were the key predictors of C-TPACK as these explained 72% of the total variance in Model 3. The addition of teaching level and CK into the model increased the variance explained by 2% (See Model 5). C-PK and teaching experience did not result in a substantial increase of the variance explained by the Models 6 and 7. Analysis of the beta values showed that all the variables had positive prediction on C-TPACK except for CK and teaching experience. The regression models also showed that the predictive power of the demographic variables of teaching level and teaching experience were less than that of TPACK constructs.

**Table 4. Stepwise regression models**

<table>
<thead>
<tr>
<th>Model</th>
<th>$B$</th>
<th>Std. Error</th>
<th>Beta</th>
<th>Sig.</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>.10</td>
<td>.19</td>
<td>n.s.</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>C-TPK</td>
<td>.92</td>
<td>.04</td>
<td>.80</td>
<td>***</td>
</tr>
<tr>
<td>2</td>
<td>(Constant)</td>
<td>-.28</td>
<td>.18</td>
<td>n.s.</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>C-TPK</td>
<td>.64</td>
<td>.05</td>
<td>.56</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>C-TK</td>
<td>.36</td>
<td>.04</td>
<td>.35</td>
<td>***</td>
</tr>
<tr>
<td>3</td>
<td>(Constant)</td>
<td>-.48</td>
<td>.18</td>
<td>***</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>C-TPK</td>
<td>.54</td>
<td>.05</td>
<td>.47</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>C-TK</td>
<td>.27</td>
<td>.04</td>
<td>.27</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>TCK</td>
<td>.22</td>
<td>.04</td>
<td>.21</td>
<td>***</td>
</tr>
<tr>
<td>4</td>
<td>(Constant)</td>
<td>-.60</td>
<td>.18</td>
<td>***</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>C-TPK</td>
<td>.56</td>
<td>.05</td>
<td>.48</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>C-TK</td>
<td>.26</td>
<td>.04</td>
<td>.26</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>TCK</td>
<td>.22</td>
<td>.04</td>
<td>.21</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Teaching level</td>
<td>.25</td>
<td>.06</td>
<td>.11</td>
<td>***</td>
</tr>
<tr>
<td>5</td>
<td>(Constant)</td>
<td>-.12</td>
<td>.22</td>
<td>n.s.</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>C-TPK</td>
<td>.56</td>
<td>.05</td>
<td>.49</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>C-TK</td>
<td>.25</td>
<td>.04</td>
<td>.25</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>TCK</td>
<td>.28</td>
<td>.05</td>
<td>.27</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Teaching level</td>
<td>.26</td>
<td>.06</td>
<td>.11</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>CK</td>
<td>-.14</td>
<td>.04</td>
<td>-.11</td>
<td>***</td>
</tr>
<tr>
<td>6</td>
<td>(Constant)</td>
<td>-.36</td>
<td>.24</td>
<td>n.s.</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>C-TPK</td>
<td>.50</td>
<td>.05</td>
<td>.43</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>C-TK</td>
<td>.27</td>
<td>.04</td>
<td>.26</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>TCK</td>
<td>.28</td>
<td>.05</td>
<td>.27</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Teaching level</td>
<td>.27</td>
<td>.06</td>
<td>.12</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>CK</td>
<td>-.20</td>
<td>.05</td>
<td>-.16</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>C-PK</td>
<td>.147</td>
<td>.06</td>
<td>.10</td>
<td>*</td>
</tr>
<tr>
<td>7</td>
<td>(Constant)</td>
<td>-.26</td>
<td>.24</td>
<td>n.s.</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>C-TPK</td>
<td>.49</td>
<td>.05</td>
<td>.42</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>C-TK</td>
<td>.25</td>
<td>.04</td>
<td>.25</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>TCK</td>
<td>.28</td>
<td>.04</td>
<td>.27</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Teaching level</td>
<td>.25</td>
<td>.06</td>
<td>.11</td>
<td>***</td>
</tr>
</tbody>
</table>

Note. ** $p < 0.01$ *$p < 0.05$. JC: junior college
Discussion

This study adds to the work of Lee and Tsai (2010), Archambault and Crippen (2009) and Graham et al. (2009) in developing instruments for measuring teachers’ TPACK perceptions for specific pedagogical contexts, which is an important area of development identified for the TPACK framework (Cox & Graham, 2009). It contributes a survey instrument for assessing Singapore teachers’ C-TPACK perceptions. Such kinds of assessment are important for understanding how teachers could develop confidence for implementing pedagogical skills that support 21st century learning (P21, 2007). The following insights about Singapore teachers’ C-TPACK perceptions have been gained from the administration of this instrument:

Gap in confidence for ICT-supported constructivist instruction

The seven TPACK constructs are comprehensive because it recognizes the interplay between content, pedagogy, and technology during ICT integration. Several implementation issues with constructivist-oriented instruction appear to be pedagogical. Gordon (2009) suggests that constructivist teaching can be easily misapplied by teachers who do not fully understand it. Windschitl (2002) further found that the constructivist epistemology tends to confront teachers with conceptual dilemmas against a more objectivist epistemology. This study, however, found that practicing teachers were fairly confident of implementing constructivist pedagogies without ICT (C-PCK). The lower mean scores of technologically-related TPACK constructs (i.e., C-TK, C-TPK, TCK, and TPACK) show that their particular struggles to be in the area of ICT-driven constructivist-oriented instruction. Yet, when considering C-TPACK, the regression model 3 shows that C-TPK, C-TK, and TCK accounted for 72% of the variance explained. This survey instrument therefore allows the specific gaps in teachers’ TPACK perceptions to be pinpointed. For these teachers, it is important to address ICT use in a constructivist context rather than constructivist instruction in general.

Importance of C-TPK

In this study, regression model 1 shows that C-TPK accounts for 64% of the variance explained whereas regression model 3 shows that the addition of TCK adds another 2% to the total variance explained by the model. These findings suggest that the teachers gave more prominence to the relationship between C-TPK and C-TPACK. In contrast, when Sahin (2011) examined teachers’ perceptions of general TPACK, they found TCK to have larger positive correlation with TPACK as compared to TPK. These findings show that teachers’ perceptions of general TPACK and pedagogy-specific TPACK are different. Their general TPACK perceptions could be more closely related to their daily teaching practices where ICT is found to be used predominantly to support information transmission activities (Gao, Choy, Wong, & Wu, 2009; Lim & Chai, 2008; Webb & Cox, 2004). Therefore, as Greenhow, Dexter, and Hughes (2008) reports, teachers tend to make ICT integration decisions by focusing on how they might represent content with technology. Windschitl (2002) emphasizes that the implementation of constructivist pedagogical practices requires new forms of expertise such as the facilitation of student learning through complex problem-based tasks and the management of knowledge constructing classroom discourse. When examining teachers’ perceptions of TPACK for e-learning facilitation, Archambault and Crippen (2009) also found TPK to be more strongly related to TPACK than TCK. Therefore, the survey results reveal that when considering pedagogy-specific TPACK, the linkages between technology and pedagogy need to be pinpointed clearly to teachers.
Need to further analyze the role of C-PCK

C-PCK was the only TPACK construct that did not predict C-TPACK. In comparison, both Sahin (2011) and Archambault and Crippen (2009) found moderate to strong positive correlations. One reason could be the item design. The stem “without using technology” may have resulted in a dissociation of these items from C-TPACK. Furthermore, Shulman’s (1986) conception of PCK, which was used to guide the development of these items, may not have been comprehensive enough to address C-PCK. In retrospect, items C-PCK1 and C-PCK2 assessed how teachers could support student thinking, which more closely describes the Constructive dimension of Jonassen et al.’s (2008) framework. More items may be needed to fully explicate the other dimensions. Therefore, these items need to be reconsidered and further validated in future studies.

Influences of demographic variables on C-TPACK are not strong

The regression analysis in this study examined several demographic variables and found that teaching level and teaching experience had significant influence on C-TPACK whereas age and gender did not. Lee and Tsai’s (2010) survey of in-service teachers with respect to TPACK for using web-based resources did not study gender effects. However, they found that older teachers with more teaching experience to be less confident of their web-based TPACK. This study found similar relationships in terms of teaching experience whereas the effects of age were not significant. These findings show that with pedagogy-specific TPACK, the influence of teaching experience needs to be carefully considered. Interestingly, the more experienced teachers perceived lower C-TPACK in this study. One explanation could be that the pedagogical practices of experienced teachers may be more strongly shaped by the school system which is still exam-driven, focusing on the dissemination of knowledge and facts (Hogan & Gopinathan, 2008). These teachers may be more established in their routine expertise, or expertise to execute fixed routines (Hatano & Inagaki, 1986) in their content area. Therefore, they may perceive greater barriers in transitioning between pedagogical approaches within the school system. This could explain the effects of teaching experience on teachers’ C-TPACK perceptions.

Lee and Tsai’s study did not compare teaching level whereas this study found that Secondary and junior college teachers to be more confident about their C-TPACK. This could be because Secondary and junior college teachers need only focus on a specific area of ICT use as they are subject specialists. On the other hand, the primary school teachers approached C-TPACK across several subjects which could explain their lower level of confidence. As there is a dearth of studies in this area, there is a need to analyze the impact of teaching levels more deeply to understand why it affects or relates to C-TPACK.

Implications for teacher professional development

Based on these findings, three implications for teacher professional development are derived:

Firstly, teacher ICT professional development programs need to help practicing teachers develop the intermediate forms of TPACK, especially C-TPK, which was perceived to have the strongest power in explaining their C-TPACK. This can be done by anchoring upon their already established C-PK to model specific pedagogical uses of ICT tools which was reflected as a gap in their TPACK perceptions. The study results also point to the need for conscious efforts to build teachers’ knowledge of ICT tools that support constructivism (C-TK) as these were also perceived to positively predict C-TPACK. On the other hand, teachers’ TCK to make constructivist use of content-based ICT tools needs to be strengthened as this will enlarge the effects of content-based tools on teachers’ C-TPACK. This, in turn may address the negative relationship between CK and C-TPACK. It is a strategy that can also strengthen the TCK of primary school teachers with respect to different subject areas, which may address the impact of teaching level on C-TPACK.

Secondly, teacher ICT professional development programs need to address the learning needs of experienced teachers. For these teachers, familiarity with teacher-directed school practices may be barriers against their confidence for C-PK and C-TPACK. These teachers need to be developed first in both C-PK and C-TK. They can then be provided with design experiences which have been proven effective for TPACK development (Koehler, Mishra, & Yahya, 2007).
Teachers are more willing to integrate ICT if it helps them to achieve their instructional goals (Zhao & Cziko, 2001). It appears that teachers in this study have yet to make this connection as no significant relationships are perceived between C-PCK and C-TPACK. Besides considering the design of C-PCK items, a third area of teacher ICT professional development could also be to facilitate teachers’ reflection on how constructivist ICT lessons can be integrated to enhance their routine pedagogies, better student performance, and fit the broader curriculum goals they need to achieve. A viable approach could be drawing upon teachers’ C-PCK by providing them with the opportunity to articulate students’ learning difficulties and misconceptions and subsequently introducing TCK that could help mediate such learning challenges, as described by Akkoc (2011).

Future directions

Several areas of future research can be considered to better understand the constructivist-oriented TPACK perceptions of practicing teachers. Firstly, this study needs to be replicated with more practicing teachers, both within and outside Singapore. The current sample of teachers is not representative of practicing teachers and validation of the instrument is still needed as split-sample analysis was not carried out in this study. The collection of qualitative data was outside the scope of this study. Besides further evaluation of C-PCK items in future studies, interviews and lesson observations can also be used to enhance understanding of teachers’ C-TPACK perceptions.

Secondly, since teaching experience and teaching levels was found to be correlated with teachers’ confidence for constructivist-oriented TPACK, more detailed studies could be made of teachers in different stages of their teaching career. Comparisons of structural equation or prediction models between these different categories of teachers could help us derive a better understanding of how the institutional context influences the adoption of constructivist-oriented TPACK. This can allow teacher professional development efforts to be better targeted at the development needs of teachers in different stages of their career, and those teaching different levels. Besides comparisons by teaching experience and teaching levels, comparisons of teachers in different subject specializations can also be made. Such studies can allow deeper insights on how constructivist-oriented TPACK can be modeled and developed in different subject areas. Finally, the survey used in this study could be validated with more practicing teachers outside Singapore. This can contribute towards the development of assessment instruments that can be used to support the evaluation of teacher development programs in ICT.

Acknowledgements

This paper refers to data from the research project “Understanding And Profiling Teachers’ Technological Pedagogical Content Knowledge (TPACK) Development Patterns” (OER12/10/KHL), funded by the Education Research Funding Programme, National Institute of Education (NIE), Nanyang Technological University, Singapore. The views expressed in this paper are the authors’ and do not necessarily represent the views of NIE.

References


An Interactive and Collaborative Approach to Teaching Cryptology

Sasa Adamovic*, Marko Sarac, Mladen Veinovic, Milan Milosavljevic and Aleksandar Jevremovic

Singidunum University, Department of Informatics and Computing, Serbia // sadamovic@singidunum.ac.rs // msarac@singidunum.ac.rs // mveinovic@singidunum.ac.rs // mmilosavljevic@singidunum.ac.rs // ajevremovic@singidunum.ac.rs

*Corresponding author

(Submitted October 4, 2012; Revised March 7, 2013; Accepted May 10, 2013)

ABSTRACT

This work proposes and describes the implementation of a novel module for the standard Cryptool educational software that provides the ability to communicate over the local computer network and the Internet. The development environment consists of the C# programming language and the open interface for Cryptool modules. The solution we propose facilitates interactive and collaborative work of students on solving cryptology problems and enables a more even learning pace across the entire group. We present and discuss practical results of our approach tested in the classroom setting during 2010 and 2011. In addition to better final grades, we have observed an increase in student interest for this area manifested in better class and lab attendance, as well as more active and creative participation. We describe two lab exercises based on the proposed solution. We evaluate the impact of our solution by means of a statistical analysis.

Keywords

Cryptology, Education, Interactive, Collaboration

Introduction

Cryptology represents an increasingly important discipline of our time. Development of modern information technologies and accompanying educational materials has led to migration of various resources to the Internet and its services. One of the most important questions that presents itself is how to protect those resources from malicious actions; in other words, how to educate security professionals capable of taking security risks and challenges?

The main problem with learning cryptology is its complexity and its foundation on complex mathematical principles and formulae. The vast majority of security solutions today does not require detailed understanding of the said mathematical apparatus, but rather a basic grasp of it, and the ability to understand its role in practical applications. The shifting of the focus from the domain of mathematical laws inherent to security solutions to the domain of practical IT applications has a significant impact on the profile of a student taking a course in this discipline. For this very reason, it becomes necessary to support the changing paradigm with appropriate approaches to learning, accompanying materials and, generally speaking, learning environment.

Our experiences with using the traditional approach to teaching cryptology have lead us to believe that a new, more interactive and collaborative approach to gaining knowledge in this area is needed. Instead of using the bottom-up approach – building from the necessary mathematical foundations towards their cryptographic application, we opted for a top-down approach – representing cryptological principles in the context of their practical application with the aim of provoking the interest for mastering the underpinning mathematical principles. In order to apply the said approach, we set out to develop educational software solution that offers the appropriate level of abstraction for solving the problems we dealt with.

This paper describes a model for interactive learning of cryptography, used in the Cryptology course at the Singidunum University during the 2010/2011 academic year. Model is based on the use of the Cryptool package with subsequently built and added network functions (details are discussed later in paper) that facilitate cooperation between students in learning more complex cryptographic functions. Paper gives examples of two exercises utilizing the model described: Caesar code and Man-in-the-middle attack. In addition, we present the evaluation of the proposed model from the standpoint of class attendance and final student grades with the discussion of improvement in comparison to the previous year.
Related work

In this section, we discuss other existing models for interactive and collaborative approaches to learning cryptography.

Rachid, Kevin and Georgios (2008) were among the first to explore the subject. In their work, they consider the fact that algorithmic animation has been the focus of intense research in many disciplines and its impact on the educational process has been marked by increasing learner autonomy. Research in this field is driven by the belief that algorithm animation can be a more effective means of instruction than manual or verbal modes of delivery. Encryption algorithms in particular, offer an interesting domain for the application of animation principles in learner/content interaction. The main challenge, however, has been how to design effective animations with a pedagogical value. This paper is concerned with the presentation of an animation of the DES algorithm that exhibits many of the features of a useful instructional material. Its pedagogical value is expressed in terms of intrinsic qualities and, in particular, the degree of interactivity and the granularity of abstraction.

Matthaus, Arno & Torben (2010) describe the development of a set of tools which allows for running large cryptanalytic jobs on a peer-to-peer (P2P) system. While P2P systems are known to scale well (BitTorrent, Skype), they are much harder to deal with than server-based systems, based on grids or cloud computing offerings. The reason to build on top of P2P, nevertheless, is to circumvent the inherent cost of any server-based solution. In this paper they show that P2P systems – while fulfilling our requirements – pose new challenges which do not exist in server-based solutions in this form. They iterate over different algorithm designs and discuss the implications of executing algorithms of these classes on a P2P system. Based on their analysis they discuss two cryptanalytic algorithms and their suitability for P2P-based computation. Additionally, they present a new fully decentralized approach for distributing the discussed algorithms in a P2P system. Their approach is specifically tailored towards scalability and different failure classes caused by malicious or unreliable peers.

Jingtao, Yiming and Lei (2009) share their experiences on the practice of interactive teaching in an information security course. The three major methods, seminar-style teaching in classroom, topic presentations and discussions, and course projects for promoting hands-on learning are described. The positive results in terms of successful learning have been witnessed on the course evaluation and the feedback from the students.

Feng, Cheng, MengXiao and YiRan (2009) consider the fact that combining the content, features and the development trend of cryptography, as well as practical experience in teaching and research, produces a teaching mode of cryptography courses based on "theory–algorithm–practice–application". According to authors, "...years of teaching and research show that the model can get better teaching results, and help train students to solve practical problems encountered in the engineering practice by using cryptography.”

In their work, Xiuli and Hongyao (2009) assume that cryptography course is designed for undergraduate students interested in this area. In order to overcome student's fear of difficulties and arouse enthusiasm in learning, authors resort to telling informative cryptography narratives to students. In addition to ensuring simplicity and ease of understanding, interactive education software was used to effectively demonstrate cryptography algorithms during the class. These made up for student lack of mathematical knowledge and application of cryptography. They encourage students to explore boldly and discuss actively on a particular subject. Through using these flexible and diverse teaching approaches, cryptography algorithms become easy to understand for the students with less mathematical background. As authors report: “undergraduate students find lots of interest in the course and are amused by the active content of the textbook.” This further motivates them to use theory and technologies of cryptography in future research projects.

Proposed solution

The first step towards improving cryptography teaching is to raise the level of interaction between students and the course material. The traditional approach to this problem consists of implementing cryptographic functions by means of classical programming languages such as Java, C or C++. The main downside to this approach is the low level of abstraction; in other words, the need for a time consuming implementation of the code that demonstrates a certain cryptographic routine. Moreover, we noticed that this option is coupled with burdensome debugging of the code due
to the fact that the expected results of cipher operations are such that it becomes difficult to determine whether an error has occurred in them or not. One also needs to take into account that the prerequisite programming skills of students within each group differ, thus resulting in different levels of ability to implement certain cryptographic routines which consequently leads to asynchronous work pace of the group.

Cryptool, on the other hand, provides higher level interaction with cryptologic components. There are not many alternatives on this field, so we selected this software as primary tool for practical laboratory classes. Lack of alternatives also limited possibilities to compare its effectiveness against other known implementations.

Using the Cryptool add-on allows for a more interactive student work with cryptographic methods compared to independent implementation of these methods by means of the above said programming languages. This tools provides the appropriate level of abstraction and visualization in working with cryptographic methods. All popular cipher algorithms are available as built-in components. One connects different routines by simply connecting their graphical representations.

Additional Cryptool functions that allow for communication over the local network and the Internet have been developed in order to facilitate student collaboration on problem solving. Also, these functions cause the group to advance at a more even pace. While utilizing the Cryptool add-in we observed considerable improvements in students results (grades, attendance, activity, creativity) which we discuss in the section devoted to evaluation of the solution.

### Adding networking capabilities to Cryptool

One of the most fundamental shortcomings of applying Cryptool in a classroom setting in a standard way is the inability to solve problems jointly, by having several students work together. The interface itself provides for a single-user work mode, whereas attempts to have two or three students solve a problem at a single workstation have failed to produce positive results. This problem becomes particularly obvious in scenarios with three participants (e.g., Man-in-the-middle attack), but also in problems where two separate sides take part in encrypting and decrypting data.

Primary contribution of this work is the addition of networking functions to Cryptool that are aimed at enabling collaborative work of students. Developing a complete Cryptool-like platform from scratch was deemed both prohibitively time consuming and laborious if one were to achieve the same level of functionality. Given the modular structure of Cryptool and the possibility for adding self-developed add-ins, the main contribution of this work is the creation of a communication capability between two running instances of this program on different computers.

### The solution model

The proposition we discuss here is implemented in the form of two novel Cryptool components which enable this tool to communicate over a network on a client-server architecture level (Figure 1). Components were developed in the C# programming language and the standard Cryptool module interface. Source code is published and available free of charge on the project web site.

The first component, **TCP server**, requires only one parameter to be entered by the user, namely the (number of the) port that is to be used by the server. Once this is done and the simulation is initiated, this component requests from the host operating system to redirect the traffic from the specified port to the component's inner structure. Incoming network traffic at the same time becomes the output of this component and is subsequently redirected towards other components – cryptographic modules within the particular scenario.

The second part of the solution is the **TCP client** component. This component requires two input parameters: IP address of the computer with a running Cryptool instance and running TCP component, and the designated port of the server in question. Input to this component is the regular expression from other components in the scenario which is sent over the network in the original form.
Using the two components we describe above makes it possible to simulate a wide array of network scenarios. In cases where a scenario might require a communication intermediary, this role is accomplished by using both components – TCP server receiving original traffic and TCP client forwarding it to the final destination (Figure 2). Other components dedicated to decrypting and alteration can be inserted in between these two.

One should keep in mind that the operation of components is not strictly limited to communication with other Cryptool instances. Instead, they may be used to connect this tool with other network systems and services. Moreover, components are in no way constrained to operation on LAN, but can be used for student collaboration over the Internet.

**Lab examples**

This portion of the paper presents two examples of lab exercises utilizing our solution. The first exercise serves as the introductory example for the course and pertains to the use of the Caesar cipher, which is one of the simplest tools employed for encryption of text messages. This exercise is carried out with students working in pairs.

The second exercise we describe is concerned with the *man-in-the-middle* type of attack. This exercise is of significantly greater complexity and requires participation of three students for collaborative learning (Lewis & Lunsford, 2010).
Caesar cipher

Caesar cipher is one of the simplest methods for encryption of text messages. It consists of shifting the letters of a particular alphabet for a previously agreed number of positions. Although it nowadays bears only a historical significance, Caesar cipher readily lends itself to demonstration of basic cryptologic concepts, be it cryptographic (algorithm and key) or cryptanalytic (attack techniques). Cryptool single-user mode provides students with the opportunity to acquaint with the Caesar cipher principle. Modules for encrypting and decrypting with Caesar cipher have been built into Cryptool.

Using proposed and developed Cryptool extension has made collaborative student work on understanding Caesar cipher possible. In this way, better insight into practical use of the cipher is gained since, in addition to the basic communication channel, students must agree on the shift (key) they will be using. In this scenario, students are arranged in pairs, where one participant is tasked with entering the open message that should be transmitted and with defining the module of the Caesar cipher in the encryption mode. Encryption result, which is the output of this module, is redirected to the TCP client, which is supplied with the IP address and the port number of the other participant's computer. This is the mechanism for exporting the cipher from the local Cryptool environment and for transfer over the local network to the remote computer environment.
The Cryptool TCP server component must be initiated on the recipient side with the aim of receiving the data sent by the encrypting party. Incoming data is redirected to the Caesar cipher module in the decryption regime. When exercise has been successful, the receiving side gets the original message. In case procedure has not been followed through correctly, the message received is not intelligible.

A necessary addition on the receiving side is the Gate component, inserted between the TCP server and the Caesar component. This component is dedicated to synchronization of recipient's and sender's environments. Namely, it delays decryption until data has arrived. Naturally, in order for the system to function properly and to establish the network connection between the parties, it is necessary to first start the environment of the recipient.

*Man-in-the-middle*

In the realm of cryptography and computer networks *Man-in-the-middle* is a complex attack scenario in which attacker actively eavesdrops on the communication channel between two persons. Attacker is entirely independent of the parties who believe they are exchanging messages over a private connection, while the entire conversation is controlled by the attacker.

Attacker may take on either a passive or active role. In case attacker is passive, he/she will be eavesdropping on the entire conversation in order to obtain information. In order for the attacker to be active, he/she must be able to intercept all messages, change their meaning and pass on the message.

In Cryptool single-user mode, students cannot simulate *Man-in-the-middle* attacks faithfully. For example, an attacker within reception range of an unencrypted Wi-Fi wireless access point can insert himself as a man-in-the-middle.

Utilization of originally developed Cryptool extensions allows for collaborative work of students on understanding the scenario of the Man-in-the-middle attack as well as devising new ways to defend from this type of attack utilizing cryptographic methods for the intended communication environment.

In this scenario, students are divided into groups of three. Two students, student A and student B, represent participants in the communication who exchange cryptologic keys by the Diffie-Hellman method (a method for exchanging cryptographic keys with symmetric key ciphers), whereas the third party is the student C assuming the attacker role (Man-in-the-middle).

In the first phase of the scenario, student A calculates information for student B based on his secret $a$ with the help of the module intended for calculation of large prime numbers. The result is passed on to the module for conversion of data types that serves to synchronize modules. Output from this module is directed to the TCP client that is provided the IP address and the port number on the workstation operated by student B. This is the sequence of steps that enables exporting of a message from local Cryptool environment and its transfer to student B over a computer network.

*Figure 6. First phase of the Diffie-Hellman algorithm on the side of student A*
In the second phase of the scenario, the Cryptool TCP server component must be started on the side of student B with the purpose of receiving data sent by student A. In this phase students A and B are oblivious to the fact that they became victims of the Man-in-the-middle attack. Received data is redirected to the module for calculation of the cryptologic key. Next, based on his secret $b$, student B calculates the information for student A and transmits the message. In case exercise is properly carried out, the receiving side, student B, gets a counterfeit message from student C playing the attacker role and posing as student A. Student B, unaware of the deception, replies to student C and establishes a joint cryptological key with the attacker.

![Figure 7. Second phase of the Diffie-Hellman algorithm on the side of student B](image)

Cryptool environment on the side of student C, the Man-in-the-Middle, includes the TCP server component that is tasked with receiving data sent by student A. Incoming data is redirected to the module for calculation of the cryptologic key. Based on the data received, a joint cryptologic key between student C and student A is determined. Next, based on his/her secret $c$ student C calculates information for student B and sends it to him. When exercise is properly executed, the recipient, student B, receives a message from student C, the attacker, posing as student A. Believing that he is communicating directly with the intended person, student B responds to student C and sets up a cryptologic key with the attacker.

![Figure 8. Third phase of the Diffie-Hellman algorithm on the side of student C](image)

When students completely understand the scenario behind the Man-in-the-middle attack they can further design a cryptographic system in which a communication between students A and B would be protected with other cryptographic algorithms.

**Evaluation**

Introduction of tools and extensions described herein into the teaching process has resulted in a good overall reception by the students and obvious positive effects. One of the most significant effects we observed is the
exchange of knowledge and insights between students. Those who were quicker to understand and adopt presented concepts had a positive impact on students that needed more time. One of the reasons for this was the wish of advanced students to have competent collaborators for scenarios that require more participants. Consequently, this helped to prevent polarization of students on the basis of different learning speeds and provided an opportunity for the group to advance faster as a whole. Also, even though attendance in lectures and exercises is not mandatory, it has risen sharply leading to an increase from 59% to 75% in more than half of the classes as compared to the previous academic year.

Significant improvement was observed in final student grades, i.e., final grades in the first three exam dates following the end of the course. During 2010/2011 distribution of student grades was approximately uniform. In year 2011 students had better grades, where a considerable portion of them attained close to the maximum number of points (Figure 9).

![Figure 9. Distribution of grades in 2010/2011 and 2011/2012](image)

We have conducted the statistical analysis of final student grades. To this end we used the T-test (Table 1). Results of the 2010 group were adopted as a control group, whereas the 2011 class – the class that has actually used the system – was chosen as the treatment group.

<table>
<thead>
<tr>
<th>Intermediate values used in calculation</th>
<th>Control Group</th>
<th>Treated Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>7.95</td>
<td>8.37</td>
</tr>
<tr>
<td>SD</td>
<td>1.43</td>
<td>1.43</td>
</tr>
<tr>
<td>SEM</td>
<td>0.16</td>
<td>0.12</td>
</tr>
<tr>
<td>N</td>
<td>84</td>
<td>138</td>
</tr>
</tbody>
</table>

- Standard error of difference: 0.198
- Degrees of freedom: 220
- T value: 2.1079
- Confidence interval: CG mean - TG mean = -0.42
  Confidence interval (95%): -0.81 to -0.03
- P value and statistical significance: P value = 0.0362
Control group mean and standard deviation were 7.95 and 1.43, respectively. In case of the treatment group, mean was 8.37 and standard deviation was 1.43. Based on the comparison of results with the corresponding values in T-table (for a statistically significant p-value of 0.0362, calculated on the basis of degrees of freedom for both groups), there exists a statistical significance of the results.

**Conclusion**

This work presents a solution for interactive and collaborative work of students on adopting the cryptology material. The solution is based on the use of Cryptool, a software package that offers a graphical interface for modeling and simulation of cryptological scenarios with the ability to use popular algorithms.

Appropriate program extensions were developed in order to facilitate collaborative work of students – namely, TCP server and TCP client. Utilization of these modules enables communication between operating environments of students working on different computers within the local network.

Deployment of the proposed solution offers an adequate level of abstraction in handling cryptographic and cryptanalytic principles and algorithms. This approach has resulted in tremendous savings of time needed to implement cryptographic procedures in classical programming languages, with particular emphasis on time saved on debugging. Focus has shifted from understanding mathematical fundamentals of cryptographic algorithms and their implementation in programming languages to understanding of architecture and weak points of complex systems. Stress was also put on using best practices to implement described mathematical systems.

In the paper, we proposed and discussed the results of our approach applied in practice. Besides better final student grades, we have observed a significant increase in student interest for the area manifested in greater dedication to lectures and exercises, as well as a more active and creative participation. In addition, better continuity of material adoption was also achieved on an individual level.

**Acknowledgments**

This work was supported by the Ministry of Science and Technological Development of Serbia through the projects TR32054, III 44006 and ON 174008. Project website: http://cryptool.singidunum.ac.rs

**References**


Adaptive vs. Fixed Domain Support in the Context of Scripted Collaborative Learning

Anastasios Karakostas* and Stavros Demetriadis
Department of Informatics, Aristotle University of Thessaloniki, Greece // akarakos@csd.auth.gr // sdemetri@csd.auth.gr

*Corresponding author

(Submitted October 14, 2012; Revised March 20, 2013; Accepted May 21, 2013)

ABSTRACT
This study focuses on how to adaptively support small groups of students during a scripted collaborative activity. Forty (40) students collaborated remotely in dyads (in lab conditions) on a task structured by a collaboration script in the domain of multimedia learning. Half of the dyads (treatment group) were supported by a domain-specific adaptive intervention in the form of reminding prompts, while the rest of the dyads (control group) were supported by an informationally equivalent fixed form of support. Our main hypothesis was that the adaptive intervention would lead to better individual and group learning outcomes compared to the fixed one. Qualitative and quantitative analyses showed that (a) students in the treatment group outperformed those in the control group in domain knowledge acquisition, (b) dyads in the treatment group accomplished tasks more efficiently than the control dyads, and (c) dyads in the treatment group enacted more solution-convergent interactions than the control dyads. Overall, this study provides evidence that by implementing techniques of adaptive domain-specific support during a collaborative activity, instructors can substantially improve learning outcomes.

KEYWORDS
Adaptive student support, Computer supported collaborative learning, Collaboration scripts

Introduction
Although collaborative learning has been proved significant for students both for social and cognitive reasons (Slavin, 1996), collaborating students usually fail to engage in productive learning interactions when left without teachers’ consistent support and scaffolding (e.g., Hewitt, 2005). Currently, issues regarding the adaptive operation of CSCL (computer-supported collaborative learning) systems attract the increasingly intense efforts of various research groups (e.g., Walker et al., 2009). These efforts advance the tradition of Adaptive Hypermedia Environments toward CSCL and expand the perspective of the field while setting innovative research agendas. In general, adaptive collaboration support techniques aim to model the major aspects of the collaborative activity and activate learner/group support interventions when needed and in the form it is needed (Soller et al., 2005). Although, there have been reported some encouraging first results (e.g., Kumar et al., 2007), there are also implementations that do not prove that such type of interventions lead to enhanced learning outcomes (e.g., Baghaei et al., 2007). Moreover, most of these systems are research prototypes that demonstrate possible system architectures or have been used to showcase their beneficial learning impact but are not widely available outside the research laboratory.

Based on the above drawbacks, we investigate if the integration of adaptive domain-specific support in a scripted collaborative activity would lead to better learning outcomes compared to a fixed support mechanism. In the following, we present (a) the theoretical background of our research, (b) the study design and results, and (c) a discussion analyzing the learning impact of the collaboration support method.

Theoretical background

Support collaboration using collaboration scripts

Collaborative learning has been proved important for students for social, cognitive and meta-cognitive reasons (Slavin, 1996). However, when students are engaged in collaborative learning they need significant support and guidance since they are rarely engaged in productive interactions such as asking each other questions or reflecting upon their knowledge (Hewitt, 2005; Liu & Tsai, 2008).
A first step toward providing the kind of student support necessary in collaborative processes has been to script the activity (Fischer et al., 2007). Scripts structure the collaborative process by defining sequences of activities, by creating roles within groups and by constraining the mode of interaction among peers or between groups (Dillenbourg & Tchounikine, 2007). Implementing CSCL scripts has been reported to result in improved learning outcomes (Fischer et al., 2007; Hernández-Leo et al., 2006). However, CSCL scripting has been criticized for its loss of flexibility (Dillenbourg & Tchounikine, 2007), and also the danger of “over-scripting” collaborative activity (Dillenbourg, 2002).

Supporting collaborative learning through adaptive/intelligent interventions

Current CSCL efforts have focused on supporting groupwork through the use of adaptive and/or intelligent systems. Adaptive and intelligent interventions tailor the collaborative learning process to the needs of the individual students or groups. The target of the adaptive/intelligent interventions varies and can be classified into 2 main categories: 1) peer interaction support (i.e., help peers to “learn to collaborate”), 2) domain knowledge support (i.e., help peers to deepen their domain understanding) (Magnisalis et al., 2011).

Peer interaction support refers to the actions taken by the system in order to help learners improve their interaction and develop domain-general knowledge and skill (Soller et al., 2005). For the purpose of this study, we focus on the second category of support. Domain-specific support refers to the actions taken by the system in order to help learners understand the domain better. This kind of support concerns the aspects of users and groups (and their activities) that have to be modeled and can be inferred or observed in system/user interaction in order to support group learning (Ayala & Yano, 1998). As domain-specific support focuses on problem-solving modeling, it involves systems that are strongly related to specific domain (Baghaei et al., 2007; Walker et al., 2009).

Four major issues emerge from the adaptive and intelligent collaboration support area: (1) systems are in an early stage of development and evaluation and relevant studies most often do not report clear learning benefits (2) the systems are strongly related to the target domain, (3) modeling students’ domain knowledge is almost always concerned with the individual, and (4) there is a lack of coherence in assessing the learning impact, since no common benchmarks have been agreed upon, making almost impossible to compare the efficiency of using different methods for supporting the same target of intervention.

Research motivation

In order to support the group learning we focus on: (a) the enhancement of student interaction and (b) the enrichment of the users’ knowledge pool. Collaboration scripts have proved to be a suitable mechanism to structure and guide student interactions. Additionally, to avoid misunderstandings, it is vital to give feedback of one's understanding and use the partner as a source for clarifications. However, the domain knowledge understanding of the group members is not always in the adequate level to foster the grounding process. In other words, even if we support students on how they should interact, we cannot be sure that they would collaborate efficiently on the specific domain. Furthermore, this lack of domain understanding may lead to the failure of collaborating partners to pool their unshared knowledge resources (Rummel & Spada, 2005). This could be fatal in a situation where the group members are mutually depending on one another's knowledge to successfully complete the group task. In order to enrich the users’ knowledge pool we could apply adaptive domain-specific collaboration techniques. Therefore, main questions emerge: can we offer adaptive support to a group of collaborating learners and is it possible for this adaptive support to be lastly cognitively beneficial for learners? Finally, will the adaptive support provide significantly enhanced learning outcomes than an informationally equivalent fixed supportive technique?

Consequently, in this study we explore whether a simple adaptive form of supportive intervention is indeed more beneficial as compared to fixed form of support. To this end, we used the “Learning Activity Management System” (LAMS) tool to implement (a) an adaptive and (b) a fixed intervention to help teammates recall important aspects of the learning material.
Method

Research objectives

The main goal of this study is to explore whether a simple adaptive form of supportive intervention is more beneficial as compared to an informationally equivalent fixed form of support when they are both provided during a scripted collaborative activity.

Experimental design

We conducted an experimental lab study comparing the two conditions: (a) students who were supported by a fixed method (control condition), and (b) students who were supported also by the adaptive prompting method (treatment condition). Furthermore, a peer-tutoring collaboration script also supported both conditions.

Instructional domain

The instructional domain of the activity was “Multimedia Learning”. This particular subdomain was part of the course “Learning theories and educational software” that the participants followed during the semester. More specifically, the domain concerned the Cognitive model of multimedia learning theory based on the Dual Coding Theory as presented in (Mayer, 2003).

Collaboration support system

The computer-based system that supported the collaboration was LAMS (Learning Activity Management System) (LAMS, 2010). LAMS is an open source licensed under GPL2 and it is basically a web-based tool for designing, managing and deploying collaborating learning activities.

The script

A two phases collaboration script orchestrated the whole activity and provided fixed form of support to the learners by guiding their collaboration. The assigned task for each phase was to provide answers to an open-ended domain questions (LAMS chat tool). These were essentially “learning questions” that provided the opportunity for structured peer interaction. However before answering each learning question, dyads were asked to discuss and agree on theory keywords that are relevant to the subject under investigation.

Overall, each one of the two phases comprised one keyword question (KQ) and one learning question (LQ). It is important to notice that students were not informed that the KQs were related with the LQs. The script also provided guidance on the roles (author and reviewer) that the students had to follow during the two LQs. One of the students was assigned the author role (responsible for introducing an initial answer) and the other one the role of reviewer (to review and propose improvements for the suggested answer). Students were then encouraged to further discuss their common answer freely, improve it, if necessary, and submit it. Afterwards, the dyad worked in a similar manner on the second phase of the script. In the second LQ peers they exchanged their roles (author, reviewer) (see Figure 1).

Treatment condition

Apart from the collaboration script, students in the treatment group were supported by one complementary method: an adaptive prompting mechanism. Dyads in the treatment mode were prompted after each keyword question (KQ) during the collaboration phase. The system was monitoring the keywords that the students provided and compared them to teacher’s keyword-based domain model (that is, seven keywords that we had pre-declared as the most important for the subject under discussion). In case some keywords were missing from the students’ dialogue, the
system responded with a relevant prompt that included information about the missing keywords. It is important to notice that the prompts were presented one by one after students had completed their discussion.

For example, a reminding prompt for the question “Think of major keywords relevant to Dual Coding Theory. Use the chat tool below to discuss and submit your list of keywords” was: “VISUAL MODEL: It seems that you have not included “Visual model” as a keyword. The theory suggests that learners organize a mental “visual model” based on perceived pictorial information”. 

The goal of the adaptive domain prompting mechanism is first to identify missing domain conceptual knowledge (as documented by analyzing the peer dialogue) and second to provide the needed information and help peers develop a more accurate, shared, individual, mental representation. In general, adaptive collaboration support techniques aim to model the major aspects of the collaborative activity (such as domain, activity structure, student/group profile, peer interactions, etc.) and activate learner/group support interventions when needed and in the form it is needed. Moreover, each statement in a group dialogue and eventually the whole group dialogue is a process of sharing the individual mental representations in order to reach common understanding. Based on this perspective, we believe that the group dialogue reflects (at least) a part of the group knowledge model based on which the adaptive support is provided. The proposed adaptive mechanism is based on each group’s domain conceptual knowledge and provides the missed and needed information. Consequently this would help group members to collaborate more efficiently (for example, during the next discussion). The triggering of the prompting technique depended on peer interaction and also the objective of the intervention was to help partners in their next task.

**Control condition**

In the control condition, after each KQ the system presented to the students the list of all keywords (seven in total) with their extended definitions that had been pre-defined by the instructor. This list of keywords is considered as a fixed support mechanism, informationally equivalent to the adaptive support mechanism, since it included all the prompts possibly presented to students in the treatment condition. Students could read the keyword definitions and then continue with the next task. In Figure 1 we present the collaborative activity and the differences of the two conditions.
Participants

The study employed 40 undergraduate informatics students (19 females) in their 3rd (out of 4) year of studies. The collaborative activity presented in this study was a required task for this course and students who successfully completed the whole procedure were awarded a bonus grade. All students were domain novices and they had never been engaged in online collaborative learning activity before.

Students were assigned to one of the two conditions (treatment and control) in the following way: we conducted a prior domain knowledge questionnaire, which included a set of eight closed-type and two open-ended question items. With this instrument, we identified the student’s basic domain knowledge (pre-test). The test was administered on a paper. Two independent raters were the raters of the test. Inter-rater reliability for the open-ended questions was high (ICC = .91).

Based on their answers, we classified students as: novice, intermediate and advanced. The next step was to assign students to dyads. Based on students’ profile, we formed 20 mildly heterogeneous dyads. In other words, all dyads consisted of students belonging in adjacent competency classes (for example, novice-intermediate or intermediate-advanced) avoiding the formation of “novice-advanced” dyads. There is evidence on the available literature indicating that mildly heterogeneous groups (regarding students’ domain knowledge) are more likely to outperform homogeneous groups at accomplishing specific goals (Wang et al. 2007). Finally, the dyads were distributed in the two conditions (treatment and control) stratified by their domain knowledge. In other words we had the same “novice-intermediate” and “intermediate-advanced” number of groups (10) in both conditions.

Procedure

The study lasted two weeks. Students first attended an introductory lecture on Monday of the first week. Then, all students were given to study the same 20 pages text-based learning material. On Monday of the second week, students were given the 30-minute prior domain knowledge pretest. The collaborative activity took place on Wednesday. There were two collaboration sessions that lasted 2 hours. The treatment group worked in the first session, and the control group in the second.

Each dyad partner was placed in different laboratories. The students in the two conditions were not informed that they would be treated differently. As the collaboration phase ended, the students individually completed the post-test in 30 minutes followed by an opinion questionnaire regarding the learning experiences. The next day, we interviewed the students from each group to record in detail their attitudes and relevant comments on the activity.

Measures

Learning outcomes

Measure 1 (individual domain learning): In order to measure the individual domain learning we conducted a post-test comprising two parts: (a) the first part included the same closed-type questions as the pre-test questionnaire. This part focused on assessing students’ acquisition on basic domain knowledge (first level of Blooms’ taxonomy) (in the following “Measure 1-A”). (b) The second part of the questionnaire included three open-ended questions, which referred to sections of the learning material that students individually studied and collaboratively worked on. This part focused on assessing the students’ understanding of the domain (second level of Blooms’ taxonomy) (in the following “Measure 1-B”).

The post-test questionnaire was developed by the authors as domain experts. The questions were additionally verified and validated by one more domain expert who was not involved in the experiment. Some ambiguous or unsuitable questions were modified, removed, altered, or arranged in a proper order. The reliability of the closed-type questionnaire was sufficiently high (Cronbach’s alpha = .842).

Measure 2 (in-task group learning): In order to measure the dyad domain learning during the task, we assessed the dyad answers to the two LQs. This is considered as “in-task group learning”, since the answers were jointly
formulated by the two partners. We transferred all the dyad answers from the log files to paper sheets. The dyad answers were assessed by two independent raters (see further below).

To avoid any biases, dyads’ and students’ paper sheets were mixed and assessed blindly by two raters, who used a 0-10 scale and followed predefined instructions on how to assess the LQs and each part of the post-test. Each student received 2 scores: (a) a score for the closed-type part of the post-test, and (b) a score for the three open-ended questions of the post-test. The mean of these two scores was used as dependent measure for individual learning. The two raters rated also each dyad answers paper sheets. The mean and the sum of these two scores was used as dependent measure for in-task group learning. As a measure of inter-rater reliability, we calculated the intraclass correlation coefficient (ICC) for the two scores (both for the LQs and the post-test). For all statistical analyses a level of significance at .05 was chosen.

Dialogue analysis

The dialogue analysis was conducted for both KQ discussions and LQ discussions. In relation to the KQ dialogues, we tried to identify for both conditions (control and treatment) the keywords that the groups had missed. For the treatment condition the missed keywords for each KQ, could be easily identified because missing keywords triggered the presentation of the relevant reminding prompt. For the control condition the missed keywords were identified through dialogue analysis, which compared the provided keywords by groups to the predefined keywords by the instructor. This analysis was conducted by the two authors and the results were the same.

Regarding the LQ dialogues, we analyzed the dialogues in order to identify the keywords that groups from both conditions missed in the KQ but included in a correct and adequate manner when formulating the respective LQ answer. To achieve that we analyzed the dialogues based on the “domain-related analysis” model proposed by Rummel and Spada (2005). This model focuses on “topics” arising within a dialogue. By “topics”, is meant short, identifiable thematic segments within a dialogue. In our study, by topics we refer to the needed keywords that a group should include in an LQ answer. Based on the model, we analyzed each topic with regard to its general relevance to the answer, the adequacy of the way in which it was discussed, the correctness of the statements and the depth of the discussion. Two coders, as before, independently coded the chat dialogues (Cohen’s kappa = .89). Disagreements were resolved through discussion. Thus, we identified the keywords that partners (in both conditions) missed in the KQ but used efficiently in their LQ answers.

Student’s opinion questionnaire

In our study the opinion questionnaire aimed to identify students’ opinions for the learning procedure in four main dimensions: (a) the adaptive prompting (domain-specific support), (b) the fixed support mechanism, (c) the main guidelines that the students had to follow in the script, (d) the whole learning experience and (e) the online collaboration environment. Naturally the first part of the questionnaire addressed only the students in the treatment mode and the second part only the students in the control group. The rest of the questionnaire was common for all the participants.

The questionnaire validity was tested through a small pilot study with 8 students who did not eventually participate in the final experiment. We administered the opinion questionnaire and conducted also short interviews after the students had worked in the environment. Based on the feedback received from the students we constructed the final form of the questionnaire. Finally, the opinion questionnaire reflects a good reliability (Cronbach alpha = .88). We statistically analyzed the students’ answers in the opinion questionnaire by calculating the mean (M) and the standard deviation (SD) for each question in each condition. Furthermore we conducted a t-test analysis to examine if the differences between the conditions are significant.

Interviews

Interviews were conducted in order to record details of how students of the different groups worked and perceived the whole activity. The interviews lasted about 15 minutes each and they were semi-structured. They focused on
students’ opinions about: (a) the activity as a whole (likes/dislikes), (b) the role of prompting (students were asked to comment on how helpful, necessary, relevant, annoying, and time consuming considered the prompts to be), (b) the role of fixed support (students were asked to comment on how helpful, necessary, relevant, annoying, and time consuming considered the fixed support to be and (c) evaluation of the learning environment in terms of usability, efficiency and workload. Students were randomly and individually interviewed.

All of the individual interviews were audio-recorded and fully transcribed for further analysis. The coding scheme analysis included four main categories and the relevant subcategories. In Table 1 we present the coding scheme.

<table>
<thead>
<tr>
<th>Category</th>
<th>Subcategories</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impact of the collaborative activity</td>
<td>-group interaction -group answers</td>
<td>Students’ statements about the benefits and shortcomings of the activity</td>
</tr>
<tr>
<td>Impact of prompting</td>
<td>-acts after the prompting -thoughts about prompts</td>
<td>Statements referring to the adaptive prompting mechanism</td>
</tr>
<tr>
<td>Impact of fixed support</td>
<td>-acts after fixed support -thoughts about fixed support</td>
<td>Statements referring to the fixed support mechanism</td>
</tr>
<tr>
<td>Impact of the script</td>
<td>-script tasks -roles</td>
<td>Statements referring to the collaboration script</td>
</tr>
<tr>
<td>The learning environment</td>
<td>-usability -efficiency</td>
<td>Statements referring to LAMS</td>
</tr>
</tbody>
</table>

The interview transcripts analysis classified students’ statements according to (a) relevance (the coding scheme categories/subcategories the statement was relevant to) and (b) attitude (the opinions or judgment expressed by students). For example, the student statement “prompts were helpful for our discussion regarding the open-ended questions” was classified as relevant to “impact of prompting” (subcategory: “acts after prompting”) and expressing the opinion of “helpful”.

In order to validate the above scheme, two independent field experts, who were not involved in any other aspect of the study, were asked to read through three transcripts (from both conditions) and to identify a category system. The generated categories were compared to the initial categories. The differences were discussed and a consensus was reached regarding the final form of the scheme. Finally two coders individually coded the interview transcripts following the above scheme. The intra-coder reliability was satisfactory (Cohen’s kappa = .90).

**Results**

**Learning outcomes statistical analysis**

We proceeded to apply parametric statistics to our data as the normality and the homogeneity of variance criteria were satisfied. T-test control was applied to students’ pre-test data and analysis of covariate (ANCOVA) to the other measures with the pre-test as covariate. The pre-test did not show any significant differences between the two group scores in either of the two questionnaire parts (part1: close type questions, part2: open ended questions) (Table 2).

<table>
<thead>
<tr>
<th>Table 2. Pre-test analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control (n=20)</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Part1 closed-type questions</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Part2 open-ended questions</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Inter-rater reliability for the second part of the post-test was high (ICC = .94). The treatment group outperformed the control group in both measures of domain basic knowledge (close type questions) and domain understanding (open ended questions). ANCOVA indicated that the difference was statistically significant (Table 3, item 1 and item 2).
### Table 3. Post test analysis

<table>
<thead>
<tr>
<th></th>
<th>Control (n=20)</th>
<th>Treatment (n=20)</th>
<th>ANCOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Measure 1-A</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(domain basic knowledge)</td>
<td>$M = 5.80$</td>
<td>$M = 7.30$</td>
<td>$F(1, 5.49), p &lt; .05$</td>
</tr>
<tr>
<td></td>
<td>$SD = 2.42$</td>
<td>$SD = 1.49$</td>
<td></td>
</tr>
<tr>
<td><strong>Measure 1-B</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(domain understanding)</td>
<td>$M = 3.19$</td>
<td>$M = 5.00$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$SD = 2.61$</td>
<td>$SD = 2.43$</td>
<td></td>
</tr>
</tbody>
</table>

Inter-rater reliability for the two LQs was also high (ICC = .905). Referring to the sum of the two LQs, ANCOVA indicated that the groups in the treatment condition outperformed the dyads in the control condition (Table 4, item 1). However, as we proceeded to interaction analysis (see below), we identified that the students’ need for support was different in the two KQs. For this reason, we further statistically analyzed the two LQs separately. In the first LQ, groups in the treatment condition achieved better scores than the groups in the control condition. This result is also statistically significant (Table 4, item 2). However, although the groups in the treatment condition answered better in the second LQ than the controlled groups, the result is not significant (Table 4, item 3).

### Table 4. In-task group learning analysis

<table>
<thead>
<tr>
<th></th>
<th>Control (n=10)</th>
<th>Treatment (n=10)</th>
<th>ANCOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Measure 2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(both LQ1 and LQ2)</td>
<td>$M = 7.88$</td>
<td>$M = 9.38$</td>
<td>$F(1, 6.88), p &lt; .05$</td>
</tr>
<tr>
<td></td>
<td>$SD = 1.96$</td>
<td>$SD = .88$</td>
<td></td>
</tr>
<tr>
<td><strong>Measure 2-1 (only LQ1)</strong></td>
<td>$M = 7.75$</td>
<td>$M = 9.5$</td>
<td>$F(1, 5.48), p &lt; .05$</td>
</tr>
<tr>
<td></td>
<td>$SD = 2.19$</td>
<td>$SD = 1.05$</td>
<td></td>
</tr>
<tr>
<td><strong>Measure 2-2 (only LQ2)</strong></td>
<td>$M = 8.00$</td>
<td>$M = 9.25$</td>
<td>$F(1, 4.34)$, n.s. ($p = .053$)</td>
</tr>
<tr>
<td></td>
<td>$SD = 1.97$</td>
<td>$SD = 1.69$</td>
<td></td>
</tr>
</tbody>
</table>

### Dialogue analysis

The analysis of the group dialogue showed that in the first KQ the treatment groups missed 18 keywords totally (same as the number of the prompts that appeared after the KQ1). Based on the analysis model, we identified that from these 18 prompts, the students used correctly and with adequacy 15 keywords in LQ1. By contrast, dyads in the control condition missed 17 prompts, but they used efficiently only 3 of them in LQ1.

Concerning the second KQ, results showed that the treatment 7 dyads missed 1 keyword each (7 keywords totally). All the dyads used that key concept efficiently during the second LQ. On the other hand, 5 control dyads missed from one keyword (5 totally) and 3 of them used it during the LQ2. In table 4 we present the results of the dialogue analysis.

### Table 5. Dialogue analysis

<table>
<thead>
<tr>
<th></th>
<th>Treatment Dyads (adaptive support)</th>
<th>Control Dyads (fixed support)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phase 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>KQ1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cases of missing keywords</td>
<td>18</td>
<td>17</td>
</tr>
<tr>
<td><strong>LQ1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cases of including key concepts in group answers after support</td>
<td>15 (out of 18)</td>
<td>3 (out of 17)</td>
</tr>
<tr>
<td><strong>Phase 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>KQ2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cases of missing keywords</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td><strong>LQ2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cases of including key concepts in group answers after support</td>
<td>7 (out of 7)</td>
<td>3 (out of 5)</td>
</tr>
</tbody>
</table>

Moreover, the dialogue analysis in LQ1 showed that the majority of the treatment dyads (8 out of 10) were focused on the subject meaning that their interactions were more solution-convergent that the dyads in the treatment group. In
contrast, the controlled dyads had the opposite behavior. The majority of the (8 out of 10) used domain aspects that were needless or irrelevant to the LQ1 solution. Finally, the interaction analysis showed no similarity with the above difference during the LQ2.

Student’s opinion questionnaire analysis

The opinion questionnaire included 9 items: (a) 3 items different for each condition concerned the support (adaptive or fixed), (b) 2 items referred to the script, (c) 2 items to the whole learning activity and (d) 2 items to the collaboration environment. In Table 6 we present the results concerning the supportive mechanisms.

<table>
<thead>
<tr>
<th>Questions</th>
<th>Answers (Likert scale 1-7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adaptive support (n=20)</td>
<td></td>
</tr>
<tr>
<td>1a  The prompts helped me to recall the relevant key concepts</td>
<td>(M=5.50, SD=1.395)</td>
</tr>
<tr>
<td>2a  The prompts help me to be more efficient when discussing the respective learning question (LQ)</td>
<td>(M=5.70, SD=1.559)</td>
</tr>
<tr>
<td>3a  The prompts were clear and precise</td>
<td>(M=6.10, SD=1.119)</td>
</tr>
<tr>
<td>Fixed support (n=20)</td>
<td></td>
</tr>
<tr>
<td>1b  I chose to see the support (keywords list) and it helped me to recall the relevant keywords</td>
<td>(M=4.15, SD=1.69)</td>
</tr>
<tr>
<td>2b  The keywords-list help me to be more efficient during the next collaborative activities</td>
<td>(M=4.60, SD=1.85)</td>
</tr>
<tr>
<td>3b  The keywords-list was clear and precise</td>
<td>(M=5.30, SD=1.750)</td>
</tr>
</tbody>
</table>

In Table 7 we present the items concerning the script, the whole learning procedure and the collaboration environment. T-test results showed that there are significant differences in students’ answers in items 6 and 7 concerning the benefits of the activity.

<table>
<thead>
<tr>
<th>Questions</th>
<th>Treatment (n=20)</th>
<th>Control (n=20)</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>4  The system guidelines for the peer roles (author, reviewer) were clear and easily understandable</td>
<td>(M=6.35)</td>
<td>(M=5.70)</td>
<td>(t=2.090)</td>
</tr>
<tr>
<td>5  I believe that I responded decent at my role</td>
<td>(M=5.95)</td>
<td>(M=5.85)</td>
<td>(t=.327)</td>
</tr>
<tr>
<td>6  The collaborative activity enhanced my domain knowledge</td>
<td>(M=5.85)</td>
<td>(M=4.80)</td>
<td>(t=2.524)</td>
</tr>
<tr>
<td>7  The collaborative activity was beneficial for me (regardless of any improvement in my domain knowledge)</td>
<td>(M=6.10)</td>
<td>(M=5.35)</td>
<td>(t=2.195)</td>
</tr>
<tr>
<td>8  The chat tool helped me to express easily my thoughts</td>
<td>(M=5.70)</td>
<td>(M=6.25)</td>
<td>(t=-1.897)</td>
</tr>
<tr>
<td>9  LAMS is a usable and pleasant environment</td>
<td>(M=6.20)</td>
<td>(M=6.60)</td>
<td>(t=-1.838)</td>
</tr>
</tbody>
</table>

Interview analysis

Students in the treatment mode were rather positive about the adaptive prompting. All of them declared that the prompts were clear, precise and understandable. They also mentioned that the prompts appeared on time. A student said: “…the prompts appeared when I could focus on them. They were short, precise with just the information that I needed in order to recall some crucial parts…” Three of the students mentioned that although they had already
studied the prompted notions, the prompts made them realize the connection between the notions and the relevant question. A student said “I was familiar with the presented information by the prompt, but I was not sure that it concerned that question until the prompt appeared…”

On the other hand, students on the control condition did not respond the same about the fixed support mechanism. 8 of them mentioned that they only checked what they had missed without further study. 5 of them studied further the keywords that they had missed and only 1 student mentioned that he/she studied all the keywords from the support. Moreover, 12 of them said that the supportive keyword list made them anxious. Some of the students mentioned that the keyword list was not interesting to them: “it was just a list with keywords and explanations… it was difficult to follow during the activity”. Another one said: “I rather prefer to task and discuss than to read a list of keywords…”

Regarding the entire activity, students from both conditions expressed a common positive view. They mentioned that it was very interesting and intriguing to discuss with a partner remotely trying to find a solution to a common problem in real time. Moreover, the strong majority of the students (n = 38) mentioned that the script roles helped them to structure and organize their discussions.

**Discussion**

It is a fact that few intelligent and adaptive collaborative support systems have been implemented and even less evaluated (Walker et al., 2009). A great deal of them is research prototypes and it is difficult not only to evaluate and determine the effect of these systems on students’ collaboration and learning, but also to deploy them in every day classroom conditions (Kumar et al., 2007). Besides, several systems are strongly domain-specific and cannot be widely used in any desired domain. Finally, there is no large-scale evidence available that proves the effectiveness of the adaptive collaboration support techniques regarding domain-specific learning outcomes (Baghaei et al., 2007).

Against the above, in this study we investigated the extent to which it is possible to effectively implement a domain-specific collaboration support strategy of dynamic format during a scripted collaborative activity. More specifically, in this study we focused on and eventually compared two kinds of support: (a) an adaptive intervention by identifying missing domain keywords in the peer discussion log file and presenting reminding prompts to partners accordingly and (b) a fixed support by giving the students the option to see all the domain keywords of a question afterwards. The adaptive method indeed resulted in improved individual and group-learning outcomes as indicated by the statistical analysis of the post-test and the LQ results respectively.

Concerning the individual learning, the first section of the post-test showed that students in the treatment mode outperformed the students in the control mode in recalling conceptual knowledge (learning at “basic” level of Bloom’s taxonomy) (table 3, item 1). Additionally, results from the second section of the post-test indicated that the treatment students used much more efficiently the conceptual domain knowledge to a new problem situation than the students in the control mode (“understanding” and “application” levels of Bloom’s taxonomy) (table 3, item 2). Also the opinion questionnaire revealed that students in the treatment mode found prompts very helpful to refresh their domain knowledge (table 6, item 1a) and believed that the prompting mechanism helped them to be more efficient during the next collaborative activities (table 6, item 2a). Finally they stated that the prompts were clear and precise enough to be understood/to be perceived (table 6, item 3a).

We also argue that the adaptive support proved more beneficial to in-task group learning compared to fixed support. Statistical analysis of the sum of the dyads scores in LQ1 and LQ2, showed that the treatment dyads achieved greater scores than the control dyads (table 4, item 1). However, although treatment dyads outperformed both in the two LQ, the results are statistically different only for the first question (table 4, item 2). This can be explained by the results of the dialogue analysis. These results showed that the main impact of the adaptive support was evident during the first phase of the script (KQ1 – LQ1) (table5, item 1 and item 2). In that phase both treatment and controlled dyads missed a large number of keywords, and as result they needed support. The assessment of dyads answers to LQ1 proved that the adaptive prompting helped the groups to answer the LQ better. On the contrary, the dyads in the second KQ missed a small number of keywords. As a result they needed less support. The statistical analysis of the dyads answers in LQ2 showed that although the treatment dyads outperformed the control dyads, the results are not significant (although it is very close to level of statistical significance) (table 4, item 3). In our opinion this can be explained by the low need for support during the second phase of the script. Moreover, interaction analysis also
revealed that the adaptive prompting helped dyads were more solution-convergent during their LQ1 discussion than the control groups. However, the two groups worked the same during the LQ2.

We believe that the improved outcomes of the treatment group can be explained by considering that the students are exposed to remedial domain-specific information right after they discuss the relevant domain issues. The adaptive mode of presentation enabled students to easily integrate missing information in the domain model they constructed by activating three key cognitive processes: “selection” (focus on relevant information), “organization” (organize new information in a coherent model) and “integration” (link it to their previous domain knowledge) (Mayer, 2003). Based also on students’ interviews we argue that one of the reasons the prompts proved beneficial is that they appeared “at the right moment”, right after the keyword discussion, simulating a human teacher intervention assessing what had been discussed. This had a positive impact on students making them feel as if they were engaged in human-to-human conversation and eventually more willing to focus on prompt information.

By contrast students in the control condition did not achieve comparable performance level as students in the treatment group. This means that the additional prompt-based support offered to treatment students was necessary for achieving the higher performance and had greater impact than the fixed support mechanism. The opinion questionnaire revealed that students in the control condition found the fixed support less helpful to refresh their domain knowledge (table 6, item 1b) and to improve the next collaborative activities (table 6, item 2b) than the treatment students (table 6, items 1b and 2b respectively). The same occurred in the (table 6, item 2a). Furthermore, the interview analysis revealed 3 major behaviours concerning the fixed support: (a) students did not give any attention at all to the support, (b) students just checked what they had missed without further study and (c) students studied only what they had missed. In contrast with the friendly tone advices of the reminding prompts in the treatment condition, the fixed support did not appeared attractive or useful to the controlled students.

Overall, we believe that the result of this study highlights an important perspective in computer-supported collaborative learning. Even when peer interactions are triggered by collaboration scripts and supported by a fixed domain-specific support, the individual domain models of the partners might not be as rich as necessary to result to an elaborated common domain model for learners. In this case, adaptive supportive mechanism (dynamic form of support) that help students “repair” their incomplete domain models can result to significantly improved learning outcomes. This perspective of collaborative learning is in line with the paradigm of sharing individual mental representations (Stahl & Hesse, 2009). The two collaborators aim to establish shared knowledge by exchanging ideas based on their internal mental representations. However, when these representations fail to meet certain criteria then the computer-based partner intervenes to the dialogue externalizing its own representations and contributing to the shared knowledge. This study shows clearly that the remedial information provided by the adaptive system can improve students’ domain models and lead to better performance in problem solving.

**Conclusion**

The limitations of this study include: (a) sample size: we acknowledge the fact that a replication study with larger sample is needed to corroborate the outcomes presented here and (b) limitations of the evaluation methods: Although the study evaluation methods showed positive learning outcomes, it needs to be further examined which part of the adaptive mechanism affected this positive impact.

We believe, also, that our study provides incentive to further explore interesting relevant questions, such as: (a) how to further automate and otherwise improve the presented method for adaptive domain-specific support by analyzing peer dialogue and (b) explore which adaptation components affect the quality of collaboration (for example using layered and decomposition evaluation frameworks (Paramythis et al., 2010)).

This study provides encouraging evidence that dynamic forms of support (as opposed to fixed forms) can be implemented during a collaborative activity and result in improved collaborative learning outcomes. The main goal of this study is to investigate specific (and simple) types of adaptive collaboration support in more detail in order to increase our knowledge of when and why adaptive collaboration support is (or not) effective. In other words, this work provides the basis for exploring the impact of more complex and thoughtful adaptive support mechanisms in the context of collaborative learning.
References


An Investigation of Reading Rate Patterns and Retrieval Outcomes of Elementary School Students with E-books

Tsung-Ho Liang and Yueh-Min Huang*

Department of Engineering Science, National Cheng Kung University, Taiwan, No. 1, University Road, Tainan City 701, Taiwan, ROC // Tel: +886-6-2757575#63336 // Fax: +886-6-2766549 // protnmike@gmail.com // huang@mail.ncku.edu.tw

*Corresponding author

(Submitted October 18, 2012; Revised April 21, 2013; Accepted May 15, 2013)

ABSTRACT
While many studies have examined the capabilities of e-books to promote children’s reading abilities, more extensive investigations of the related reading patterns and outcomes seem necessary before this technology is widely adopted. This study used a reading rate tracking technique to collect the reading rate data from 24 sixth-graders participating in six reading tasks. After each reading task, the students were asked to complete a retrieval test in order to assess their reading outcomes. Using a two-stage cluster analysis method, this study identified two different reading rate patterns, the Coherent and Fluctuant Readers, from 2,820 on-reading records collected by a reading rate tracking technique. The characteristics of the patterns not only represent the changes in specific reading rate related to different groups, but can also be regarded as the extrinsic indications of various reading actions or navigation strategies, and are thus somewhat associated with the retrieval outcomes. Moreover, this study also found that e-book reading is plausibly able to promote the student’s retrieval as compared with reading printed books. Future research can examine how these patterns result in different outcomes, and thus further help teachers as they seek to guide students’ reading of e-books.

Keywords
Reading rate, Retrieval outcome, Reading e-books, Cognitive process, Elementary school

Introduction
As e-books on various electronic reading devices are currently transforming the way people read, a number of researchers have investigated how to exploit this new technology for educational purposes (Bierman, Ortega, & Rupp-Serrano, 2010; Grimshaw, Dungworth, McKnight, & Morris, 2007; Woody, Daniel, & Baker, 2010). Several studies have demonstrated that the capability of e-books to incorporate audio-visual materials is able to provide on-demand multimedia materials to support personalized reading (e.g., de Jong & Bus, 2004; Grimshaw et al., 2007; Korat & Shamir, 2007, 2008). For example, Wood (2005) reported that the use of a computer-based storybook was associated with gains in rhyme detection ability among beginning readers.

However, some studies found that both teachers and students still prefer using printed books (Bierman et al., 2010; Woody et al., 2010), and this echoes the view of Coyle (2008) that the e-book industry is currently more focused on how to render the printed books electronically, rather than developing and using new technologies that actually facilitate reading. Woody et al. (2010) suggested that the designers of e-books should consider how to develop them for optimal use, and thus they need to further explore the various user experiences that occur when reading e-books. This is especially true with regard to children, who seem to prefer screen-based reading and are generally very capable at using digital devices (Huang, Liang, Su, & Chen, 2012; Liu, 2005). Despite the many studies that examine the e-book reading of undergraduate students, little effort has been made to explore children’s use of e-books (Salmerón & García, 2011), including the related reading traits, reading preferences and cognitive processes. It is thus necessary to conduct more intensive studies with regard to the use of e-books with children in educational contexts.

Reading patterns are a major criterion for assessing a student’s overall reading situation, and may be linked to specific cognitive processes and reading outcomes (Cole et al., 2011). In practice, if a student is frustrated when reading, some form of encouraging stimulation or adaptive guidance should be introduced at that point. However, observing and understanding an individual student’s reading patterns is difficult in a traditional classroom, where such adaptive learning guidance needs to rely on the teacher’s professional judgment, based on their pedagogical knowledge and experience.
Since reading is carried out using the eyes, several studies (Erickson et al., 2011; Jukka, 2010; Liu & Shen, 2011; Marisa, 2011; van Gog & Scheiter, 2010) use eye-tracking techniques to record the location of a reader’s gaze and the fixation duration to examine how information is acquired from texts (Cole et al., 2011). For example, Nielsen (2006) identified an F-shaped reading pattern by analyzing the users’ eye-tracking data when reading web content. Cole et al. (2011) also adopted this technique to undertake an investigation of the eye movement reading patterns that occurred during information search tasks. Unfortunately, such techniques cannot be applied to ordinary classroom practice, and an alternative approach is to use an e-book reading platform along with an efficient analysis procedure, as this can provide a low cost and relatively unobtrusive way to examine the reading processes of individual students (Huang et al., 2012).

Many studies (Carver, 1990; Duggan & Payne, 2009; Dyson & Haselgrove, 2000; Fraser, 2007; McLay, 2007; Rasinski, 2000; Stroud & Henderson, 1943) state that the reading rate (the words read per minute, wpm) is a useful indicator that reflects an individual’s cognitive processing of an incoming text. For example, Fraser (2007) compared the reading rate and performance of two groups of Mandarin speakers as they carried out five tasks, and found that the varied substantial differences in the reading rates existed between the groups for all tasks. The reading rate can also be used as a tool to assess students’ reading performance (Rasinski, 1999, 2000), and has been shown to be positively correlated with reading comprehension (Joshi & Aaron, 2000). Therefore, by a reading rate tracking technique embedded in an interactive e-book learning system (Huang et al., 2012), reading rate patterns can be extracted from the students’ reading process profiles so as to provide details of specific changes reading rate for various students’ reading styles. Just as Cole et al. (2011) investigated information acquisition and cognitive processing by modeling students’ information search behaviors, it is necessary to understand reading rate patterns and the related reading outcomes (e.g., retrieval performance) in order to examine readers’ information gain and cognitive processes. Moreover, this approach may be useful for enabling teachers to rapidly perceive the problems that individuals face when reading, and this offer appropriate assistance.

To this end, this study aimed to examine students’ reading rate patterns when reading e-books, and further identify the associations among these patterns and different retrieval outcomes. In addition, the difference in retrieval outcomes between reading e-books and printed books was also examined in order to consider the effects of these two different conditions on reading performance.

**Background and related work**

**Identifying reading processes from a cognitive perspective**

Johnston, Barnes, and Desrochers (2008) claimed that the cognitive models involved in the comprehension-related skills can depict how a fluent reader accesses textual information based on individual knowledge and reading goals. They thus suggested that these models can be used to identify readers’ comprehension skills and associated cognitive functions, such as the use of a working memory model, to reveal the typical development of individual differences in reading comprehension.

The cognitive theory of multimedia learning illustrates how three kinds of materials (pictures, spoken, and printed words) are processed within three kinds of memory storage (sensory, working, and long-term memories), which represent different cognitive processes in the human information-processing system (Mayer, 2005). Since a variety of multimedia materials serve as the reading resources in e-books, it stands to reason that the e-book reading process can be regarded, by cognitive theory, as a multimedia learning process. However, as yet few studies have shed light on the e-book reading process from cognitive perspectives, and this is one motivation for the current research.

Taking the simplest reading condition as an example, Figure 1 shows that printed words are presented visually and initially processed through the reader’s eyes (step a). The reader may then attend to some of the incoming words (step b) and bring them into working memory. The working memory serves as a mental workspace where information is retrieved and integrated with incoming text. Subsequently, by mentally pronouncing the written words the reader can get them into the auditory/verbal channel (step c). Next, the active cognitive process can take place to build the words into a coherent mental structure known as a “verbal model”, as shown in step d in Figure 1. Next, the
reader may apply prior knowledge in the long-term memory to guide the process of integrating knowledge in working memory (steps e and f). Finally, after new knowledge is constructed in the working memory, it is then stored in the long-term memory (step g), where it will serve as prior knowledge to support subsequent new learning. Both the reading rates (step a) and retrieval outcomes (step f) were measured in this study to identify the students’ patterns and reading outcomes.

**Figure 1.** The cognitive processes associated with printed words

*Encoding and retrieval procedures in memory*

Effective reading requires that readers retain information in their working and short-term memories at various levels of comprehension (Ecker, Lewandowsky, Oberauer, & Chee, 2010; Nation & Cocksey, 2009; Swanson & Howell, 2001). Young learners need to learn how to decode a large number of words to master the reading process, and their limited word knowledge, along with other factors, like motivation and self-confidence (Netten, Droop, & Verhoeven, 2011), may result in great individual differences in abilities, making the examination of their reading processes even more challenging.

Cognitive psychology considers that three procedures are essential for complete memorization to occur, namely encoding, storage, and retrieval (Rajaram & Barber, 2008). A reading process, as mentioned above, can be regarded as a series of cognitive actions in three procedures. It begins with the encoding procedure of constructing new knowledge into the working memory through the related cognitive processes (in Figure 1, steps a to f), and it is then stored in the long-term memory as prior knowledge, as in step g. This prior knowledge may be retrieved later, as in the retrieval outcomes considered in this study, to support the integration of new knowledge in step f. Therefore, realizing the implications of various retrieval outcomes is critical to better understanding the encoding and retrieval procedures for an individual student, and this is why the retrieval outcomes were assessed for the students’ reading performance in this study.

*Exploring reading rate patterns by the tracking technique*

With regard to the cognitive processes shown in Figure 1, the reading rate (measured from step a) is plausibly associated with encoding and retrieval efficiency, and differs depending on a reader’s purpose, such as comprehending or memorizing the text. For example, a typical optimum reading rate (reading rate) is 260-300 wpm for a university student (Fraser, 2007). However, Carver (1983) stated that a student’s reading rate will vary slightly depending on the reading materials used or tasks undertaken, while McLay (2007) reviewed the reading rate research and claimed that measuring this rate is difficult, because it changes with how fast a text is being processed and comprehended.

As shown in Table 1, based on the reading rates and statuses associated with various reading behaviors in the literature, it should be possible to use the reading rate as an indicator to identify various behaviors (Huang et al., 2012), and this is the basic concept underlying the data collection process used in this work.
Table 1. The reading rates and statuses associated with reading behaviors in the literature

<table>
<thead>
<tr>
<th>Reading status</th>
<th>Reading rate (wpm)</th>
<th>Reading behavior</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>On-reading</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slowing</td>
<td>0-1,000</td>
<td>Excessively slow</td>
<td>(Harris &amp; Sipay, 1990; Rasinski, 2000; Walczyk, Marsiglia, Bryan, &amp; Naquin, 2001)</td>
</tr>
<tr>
<td></td>
<td>&lt; 50</td>
<td>Inefficient reading</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Disfluent</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Labored</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inexpressive</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unenthusiastic rendering</td>
<td></td>
</tr>
<tr>
<td>Memorizing</td>
<td>50-100</td>
<td>Sustained attention</td>
<td>(Carver, 1977, 1990; Duggan &amp; Payne, 2009; Fraser, 2007; Gillett &amp; Temple, 1986; Harris &amp; Sipay, 1990; Z. M. Liu, 2005; Z. M. Liu &amp; Huang, 2008; Rasinski, 1999; Reader &amp; Payne, 2007; Reading, 2012; Stroud &amp; Henderson, 1943)</td>
</tr>
<tr>
<td>Learning</td>
<td>100-200</td>
<td>In-depth reading</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oral reading</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Concentrated reading</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Annotation (highlight)</td>
<td></td>
</tr>
<tr>
<td>Rauding</td>
<td>200-400</td>
<td>Silent reading※</td>
<td></td>
</tr>
<tr>
<td>Skimming</td>
<td>400-700</td>
<td>Keyword spotting</td>
<td></td>
</tr>
<tr>
<td>Scanning</td>
<td>700-1,000</td>
<td>Reading selectively</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Browsing and scanning</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-linear reading</td>
<td></td>
</tr>
<tr>
<td>Off-reading</td>
<td>≥ 1,000</td>
<td>Flip page</td>
<td>(Carver, 1977, 1984; Harris &amp; Sipay, 1990)</td>
</tr>
<tr>
<td>Flipping</td>
<td>≥ 1,000</td>
<td>Flip page</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Glance and glimpse</td>
<td></td>
</tr>
</tbody>
</table>

Note. “※” represents a reading behavior examined in this study.

Research design

The design of the reading tasks

This study investigated the reading rate patterns that occurred while elementary school students were reading e-books, with a particular focus on the general reading situations occurring in their daily learning activities. In practice, reading is usually carried out in a rauding status (Carver, 1977) at an individual’s preferred reading rate, at which they can effectively assimilate the ideas across sentences at a pace that their cognitive speed allows (Fraser, 2007).

This study examined various possible reading events, and three different reading durations (15, 30 and 45 minutes) were examined based on the length of a regular class period (40 minutes) at elementary schools in Taiwan. Two different reading conditions, e-books and printed books, were used to investigate whether there were any differences in the related retrieval outcomes. A unique sample group with a repeated measures design (also known as a within-subjects design) was adopted in this work to compare the same measure under two different reading conditions.
Therefore, the printed book reading tasks carried out in this study (as shown in Figure 2a) were designed as the comparative condition, in comparison with the e-book reading one (as shown in Figure 2b).

Participants and reading materials

This study recruited 24 sixth-graders (12 males and 12 females) from an elementary school in Taiwan, who all volunteered to participate in the experiment, with their parents’ consent also being obtained. Before the experiment, the researcher taught the students how to use the e-book learning system to read e-books, and also reminded them of the procedure they needed to follow when doing the activity. As shown in Figure 3, the students were asked to silently and normally read three e-books and three printed books in three weeks (one e-book and one printed book per week). Both sets of e-books and printed books had three volumes that were assigned to be read in 15, 30, and 45 minutes, with longer books being given more time. An observer was present throughout the reading process, so that students could ask for help if necessary.

The six books (three e-books and three printed books) contained general science information written in Chinese. As shown in Table 2, the amount of words ranged from 3,633 to 9,180, and the average words per sentence ranged from 13 to 15. Based on Johnson, Kress, and Pikulski (1987) recommendation that word recognition accuracy rates of 95% or better are needed for successful instructional reading, this study selected texts in which from 96.54% to 99.48% of the words appeared in a list of 5,021 common Chinese words produced by the Taiwanese Ministry of Education (2000), and thus most should have been familiar to these students.

Based on Carver (1990) suggestion that texts should be easy to read and the participants should be instructed to read normally, texts at this level were used to ensure that all the students could recognize the words, preventing any ceiling effects for both worse and better decoders.

![Figure 3. The procedure used for the reading tasks](image)

| Table 2. The textual analysis of the six books used in this study |
|------------------|--------------|--------|--------|--------|
| Materials        | Textual analysis | 15 minutes | 30 minutes | 45 minutes |
| E-books          | total number of words | 3,640 | 5,475 | 9,180 |
|                  | proportion of common words | 99.09% | 99.20% | 98.99% |
|                  | average words per sentence | 15 | 14 | 14 |
| Printed books    | total number of words | 3,633 | 5,374 | 9,177 |
|                  | proportion of common words | 99.48% | 98.94% | 96.54% |
|                  | average words per sentence | 15 | 15 | 13 |

Identifying the reading rate patterns

The different reading conditions (by e-book vs. printed book) and reading durations (15, 30 and 45 minutes) were considered as the different reading conditions to examine the retrieval outcomes for e-books and printed books. In
the e-book reading condition, the reading rate tracking function embedded in the system was used to simultaneously record the reading rate in the reading process profile of every student (as shown in Figure 4). The printed book reading condition was used for the comparative reading tasks, since previous studies (Reinking, 1997, 2005; Salomon, 1979) also used printed texts, and they should be examined to ensure that reading e-books did not lead to poorer outcomes. Since reading rate tracking was only embedded in the e-book system, and thus was not available in the printed book reading condition, the reading rates were not recorded for the latter.

The three average reading rates of the e-book tasks were used as variables to analyze the different reading patterns, and these are defined as follows:

Rate_15: the average reading rate of the 15-minute reading task.
Rate_30: the average reading rate of the 30-minute reading task.
Rate_45: the average reading rate of the 45-minute reading task.

A two-stage cluster analysis method (Chiu et al., 2010; Liang, Wang, & Huang, 2008) was applied to identify the reading rate patterns based on these three variables. An independent samples t-test method was used to examine the significance among the patterns.

The outcomes of the retrieval tasks

Upon completing every reading task, the students were asked to do a 10-minute retrieval test, as shown in Figure 3. Specifically, they were asked to freely recall a portion of the information they had just learned (Carver, 1990; Fraser, 2007), and the outcomes were regarded as their individual reading performance. This is because if a student did not completely read and understand the text, then the new knowledge might have been weakly stored in their long-term memory, or even lost from the working memory, based on the theoretical background mentioned above.

In the retrieval tests, the students were asked to write 10 sentences based on the scientific knowledge they had just read, following the original text as accurately as possible. The retrieval outcomes were then measured according to the accuracy of each sentence using five discrete levels (the rating criteria were: four points for completely the same meaning as the text, three points for almost the same, two points for half the same, one point for a little the same, and zero points for no answer or completely wrong). Two raters assessed the sentences based on these rating criteria. The Cronbach’s alpha coefficient of their preliminary rating results was .87 (ranging from .79 to .93), showing a good fit for the reliability measure. The raters were then asked to review the results together and reach a consensus.

In a sense, the retrieval score represents the degree to which an individual is able to correctly retrieve information from their prior knowledge in the long-term memory (Hunt, 2008). The paired samples t-test method was used to...
compare the retrieval outcomes for reading e-books and printed books. The independent samples t-test method was carried out to examine the significance of the retrieval outcomes of the reading rate patterns found in this study.

Results

Two reading rate patterns: The coherent (Cluster 1) and fluctuant readers (Cluster 2)

A total of 5,919 records were collected from the students’ reading process profiles when reading e-books, and divided into 2,820 on-reading and 3,099 off-reading records by the system. This is based on Carver’s (1977, 1984) observation that it is impossible for an individual to read a passage at 1,000 wpm or more, and thus if the rates were greater than or equal to this they were classified as “off-reading”, while those less than 1,000 wpm, were classified as “on-reading.” This study thus only considered the on-reading records when calculating the reading rate, and the off-reading records, i.e., page flipping behavior, were eliminated from the analysis.

The two-stage cluster analysis method was adopted to classify the students into two clusters. Discriminant analysis was used to verify the validity and stability of the clustering. The results show that 95.8% of the original and 87.5% of cross-validated grouped cases were correctly classified, which demonstrates that the clustering results were robust. The final cluster centroids are shown in Table 3.

Table 3. Cluster centroids from the K-means cluster analysis results

<table>
<thead>
<tr>
<th>Variables</th>
<th>Cluster 1 (n=16)</th>
<th></th>
<th>Cluster 2 (n=8)</th>
<th></th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate_15</td>
<td>-0.52</td>
<td></td>
<td>1.04</td>
<td></td>
<td>28.63***</td>
</tr>
<tr>
<td>Rate_30</td>
<td>-0.50</td>
<td></td>
<td>1.01</td>
<td></td>
<td>24.53***</td>
</tr>
<tr>
<td>Rate_45</td>
<td>-0.38</td>
<td></td>
<td>0.76</td>
<td></td>
<td>9.61**</td>
</tr>
</tbody>
</table>

Note. * p < .05, ** p < .01, *** p < .001. The cluster descriptors are based on standardized scores (mean value = 0 and standard deviation = 1). “n” represents the number of students in related clusters.

In order to understand the fluctuations in reading rate that occurred throughout the reading process, Table 4 shows the descriptive statistics of the changes in reading rate every 5 minutes for different student clusters. The t-test results for the three variables (as seen in the Average row in Table 4) show a significant difference between the two clusters, and this is also found in the averages of the reading rate distributions between clusters in all three reading tasks, as seen in the Average row of Table 5.

Table 4. The descriptive statistics and t-test results of changes in reading rate for different student clusters

<table>
<thead>
<tr>
<th>Variables</th>
<th>15-minute reading task (n = 603)</th>
<th></th>
<th>30-minute reading task (n = 1,007)</th>
<th></th>
<th>45-minute reading task (n = 1,210)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>Cluster 1 m (s.d.)</td>
<td>Cluster 2 m (s.d.)</td>
<td>t</td>
<td>Cluster 1 m (s.d.)</td>
<td>Cluster 2 m (s.d.)</td>
</tr>
<tr>
<td>5</td>
<td>282.25 (57.73)</td>
<td>543.58 (122.61)</td>
<td>-5.72***</td>
<td>325.38 (124.64)</td>
<td>446.39 (106.44) -2.35 *</td>
</tr>
<tr>
<td>10</td>
<td>312.85 (86.52)</td>
<td>462.51 (114.56)</td>
<td>-3.59 **</td>
<td>346.69 (115.11)</td>
<td>454.08 (91.34) -2.29 *</td>
</tr>
<tr>
<td>15</td>
<td>377.47 (109.63)</td>
<td>538.17 (140.91)</td>
<td>-3.08 **</td>
<td>321.28 (104.84)</td>
<td>443.07 (105.07) -2.68 *</td>
</tr>
<tr>
<td>20</td>
<td>334.79 (118.51)</td>
<td>460.94 (136.34)</td>
<td>-2.34 *</td>
<td>337.89 (117.53)</td>
<td>537.98 (122.86) -2.76 *</td>
</tr>
<tr>
<td>25</td>
<td>353.06 (80.15)</td>
<td>513.55 (118.13)</td>
<td>-3.95 **</td>
<td>306.00 (145.02)</td>
<td>360.67 (196.98) -0.77</td>
</tr>
<tr>
<td>30</td>
<td>328.55 (176.36)</td>
<td>532.42 (156.87)</td>
<td>-2.76 **</td>
<td>287.39 (163.67)</td>
<td>412.03 (164.83) -1.75</td>
</tr>
<tr>
<td>35</td>
<td></td>
<td></td>
<td></td>
<td>235.52 (222.68)</td>
<td>386.83 (244.67) -1.52</td>
</tr>
<tr>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td>211.01 (209.12)</td>
<td>313.74 (311.60) -0.96</td>
</tr>
<tr>
<td>45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. * p < .05, ** p < .01, *** p < .001. “n” represents the number of selected records. “Average” represents the average reading rate of the related reading task.
Cluster 1: Coherent readers

Sixteen students (66.67%) were classified into Cluster 1, namely Coherent Readers, as the changes in their reading rates were more consistent with the average value (ranging from 303.34 to 334.96 wpm) than those of the students in Cluster 2.

Cluster 2: Fluctuant readers

Eight students (33.33%) were classified as Fluctuant Readers, since their average reading rates (ranging from 423.08 to 514.76 wpm) were more varied and higher than those in Cluster 1. In addition, these students read significantly faster and produced more reading rate records than those in Cluster 1, as shown in Tables 4 and 5.

Table 5. The descriptive statistics and t-test results of the reading rate distributions for different student clusters

<table>
<thead>
<tr>
<th>Reading rate range</th>
<th>15-minute reading task (n=963)</th>
<th>30-minute reading task (n=2,003)</th>
<th>45-minute reading task (n=2,953)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cluster 1 m (s.d.)</td>
<td>Cluster 2 m (s.d.)</td>
<td>t</td>
</tr>
<tr>
<td>On-reading</td>
<td>22.75 (4.04)</td>
<td>29.88 (4.58)</td>
<td>-3.90 ***</td>
</tr>
<tr>
<td>Off-reading</td>
<td>9.94 (10.58)</td>
<td>25.13 (23.22)</td>
<td>-1.76</td>
</tr>
<tr>
<td>Average</td>
<td>32.69 (12.94)</td>
<td>55.00 (21.41)</td>
<td>-3.20 **</td>
</tr>
</tbody>
</table>

Note. *p < .05, **p < .01, ***p < .001. “n” represents the number of selected records. “Average” represents the average reading rate records of the related reading task.

Retrieval outcomes between different patterns

This study conducted a paired sample t-test to find out the paired differences in retrieval outcomes between e-books and printed books. The only significant difference was found (t = -2.68, df = 23, p < .05) in the retrieval score for the 45-minute reading task. According to the paired sample statistics, the average of retrieval scores for the e-book (m = 11.96) in the 45-minute reading task was lower than that of the printed book (m = 14.58) was higher than that of the printed book (m = 11.96) in the 45-minute reading task. In other words, reading e-books for a longer period, in this case 45 minutes, seemed to lead to better outcomes with regard to retrieval than was seen with the printed books.

An independent samples t-test was also conducted to compare the retrieval outcomes between the Coherent and Fluctuant Readers. As shown in Table 6, no significant differences were found between these two groups, which shows that different reading rate patterns may have the similar retrieval outcomes.

Table 6. The descriptive statistics and t-test results for the retrieval outcomes

<table>
<thead>
<tr>
<th>Materials</th>
<th>Reading time</th>
<th>Coherent Readers (n = 16)</th>
<th>Fluctuant Readers (n = 8)</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>m (s.d.)</td>
<td>m (s.d.)</td>
<td>m (s.d.)</td>
<td></td>
</tr>
<tr>
<td>Printed books</td>
<td>15</td>
<td>15.38 (6.01)</td>
<td>12.38 (7.23)</td>
<td>1.08</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>15.44 (4.97)</td>
<td>19.75 (4.65)</td>
<td>-2.05</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>11.63 (5.45)</td>
<td>12.63 (4.98)</td>
<td>-0.44</td>
</tr>
<tr>
<td>E-books</td>
<td>15</td>
<td>14.31 (5.61)</td>
<td>14.75 (4.13)</td>
<td>-0.20</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>17.88 (6.61)</td>
<td>16.38 (5.85)</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>13.44 (6.63)</td>
<td>16.88 (5.08)</td>
<td>-1.28</td>
</tr>
</tbody>
</table>

Note. *p < .05. “n” represents the amount of students in the related group.

As shown in Table 7, the inter-correlation results between the retrieval outcomes and the reading rates revealed that, overall, the latter do not appear to be closely associated with the former whole. Only the reading rate of the 30-minute reading task was moderately associated with the retrieval outcomes for all students, and this relationship was found for the Coherent Readers but not the Fluctuant ones.
Table 7. The inter-correlations for retrieval outcomes and reading rates

<table>
<thead>
<tr>
<th>Measure</th>
<th>All Students</th>
<th>Coherent Readers</th>
<th>Fluctuant Readers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rate_15</td>
<td>Rate_30</td>
<td>Rate_45</td>
</tr>
<tr>
<td>Retrieval outcomes</td>
<td>.350</td>
<td>.462 *</td>
<td>.203</td>
</tr>
</tbody>
</table>

Note. * Correlation is significant at the .05 level (2-tailed).

Discussion

Reading rate patterns

This study used the reading rate tracking technique to identify the individual reading rates from the students’ reading processes, overcoming the difficulty of measuring the reading rate that is encountered with printed books (McLay, 2007), and also doing so unobtrusively. Based on the findings of this work, the reading rate, as noted by many other researchers, is a very effective indicator to identify reading behavior, and can be used to distinguish different reading patterns from reading process profiles. Our contribution compared those studies that used eye-tracking techniques to identify reading patterns, is not only easier to use in educational practice, but also provides more details on the reading characteristics of individual learners when using e-books.

For example, as shown in Tables 4 and 5, the Fluctuant Readers read e-books at high speeds, either skimming or scanning, as described in Table 1. In contrast, Coherent Readers maintained a steady reading rate, usually the reading status with less off-reading records, and thus were more immersed more in the texts than the Fluctuant Readers. The reading rate patterns revealed in this study can be used to identify different types of reading behavior with e-books, which is useful in modeling students’ reading behavior and investigating knowledge acquisition and cognitive processing when using e-books. Students with different reading patterns might use different reading rates for different tasks in order to assimilate the ideas across sentences at a pace that their cognitive speed allows (Fraser, 2007). However, more work is needed to uncover the sophisticated processes at play here.

Retrieval outcomes: Classifying reading performance from a cognitive perspective

A retrieval test can measure what the students have remembered and what they have learned (Rajaram & Barber, 2008), and this work examined the difference in retrieval outcomes between using the e-books and printed ones. Based on the finding that the retrieval outcomes when reading e-books for a longer period (i.e., the 45-minute reading task) were better than those when reading the printed ones for all the participants, it can be concluded that e-books are able to lead to better retrieval outcomes compared to printed ones, although this effect may have been too subtle to notice in this study.

How are the reading rate patterns associated with retrieval outcomes?

According to the retrieval outcomes shown in Table 6, no significant differences in retrieval performances were seen among all the reading tasks, and thus similar retrieval outcomes were associated with different reading rate patterns.

The reading rate, which is associated with cognitive speed (Fraser, 2007) and reading performance (Rasinski, 2000) in the literature, represents an individual’s reading behavior (Huang, Liang, & Chiu, in press; Huang et al., 2012). However, as the results in Table 7 show that significant correlations were found between the retrieval outcomes and the reading rates in the 15- and 30-minute reading tasks for Coherent Readers, but not for the Fluctuant Readers, which reveals that the reading rates may not be closely associated with the retrieval outcomes for all the students examined in this work. This may be because Fluctuant Readers read significantly faster and produced more reading rate records than the Coherent Readers, which seems to imply that different reading strategies were adopted for different reading rate patterns.

Therefore, reading rate patterns can thus be considered as reading actions or navigation strategies (Lawless, Mills, & Brown, 2002; Salmerón & García, 2011), and viewed as different styles of encoding information in the memory. In
In this study, all students were asked to complete the same reading tasks and had similar retrieval outcomes. Based on this finding, we infer that the students’ comprehension abilities, which enabled them to comprehend the text in their working-memories, were also similar, even though they used the different reading actions or navigation strategies mentioned above. In order to confirm this, two comprehension tests were conducted after the experiment, and no significant difference in the average comprehension scores was found between the students with different reading patterns ($t = -0.586, df = 22, p > .05$), supporting this inference.

However, the relatively short experimental activities used in this work, without any instruction, may have been a limitation with regard to enhancing the students’ retrieval outcomes, echoing the claim in Reinking (2005) that simply exposing students to digital reading does not guarantee instant improvements or changes in general comprehension ability. Reading individually may affect the students’ comprehension strategies in the intrinsic cognitive process, as shown in Figure 1, and then affect their extrinsic reading behaviors, such as their reading rates. For this reason, this study classified these adaptive reading rates as varied reading patterns so as to represent the students’ extrinsic reading behaviors, and identified the links between the reading rate patterns and retrieval outcomes. However, factors such as comprehension ability (Reinking, 2005) and reading strategy (Park & Kim, 2011) may dominate reading performance with regard to cognitive memory, and future research should further explore how these affect e-book reading performance (Cole et al., 2011; Salmerón & García, 2011).

**Conclusion**

This study examined elementary school students’ reading rate patterns when reading e-books, and investigated how the various patterns are associated with different retrieval outcomes from a cognitive perspective. Students with two different reading patterns, classified as *Coherent* and *Fluctuant Readers*, were identified by applying the two-stage cluster analysis method to the reading rate data. The characteristics of the reading rate patterns show that the students adopted specific reading rates when undertaking the e-book reading tasks.

Based on our findings, the reading rate is a very useful indicator for identifying reading behavior, and it can be used to differentiate various reading rate patterns in the reading process profiles. This study also examined the students’ retrieval outcomes, providing more details of individual retrieval performance from a cognitive memory perspective, which can be used to assess what the students have remembered. In addition, this study also found that e-book reading is slightly better with regard to retrieval compared with reading printed books, although the difference is very slight.

Finally, this study suggests that reading rate patterns should be regarded as the extrinsic representation of reading actions or navigation strategies, and that students’ reading rates seem to somewhat associated with their retrieval outcomes. However, future studies should collect data from multiple e-book reading contexts, use larger sample sizes, and gather more empirical evidence in order to confirm the findings of this work.

E-books are now becoming more popular, and have started to be applied in educational contexts. By obtaining more empirical evidence about reading patterns and retrieval outcomes, adaptive reading guidance can be derived to support the cognitive processes of students. Moreover, a ubiquitous learning context thus was properly formed with e-book reader portability, in which students may gain authentic knowledge by interacting with learning material and the real environment and further achieve meaningful learning (Huang, Chiu, Liu, & Chen, 2011; Huang, Huang, Huang, & Lin, 2012). The future design of e-book reading system could be designed in consideration of fostering metacognitive capability by taking mobility advantage of e-book (Liu, Huang, Kinshuk, & Wen, 2013). More studies in examining how to efficiently acquire information from various media in e-books is also critical, especially as digital natives will soon be the majority of learners, and thus such reading technologies will be ever more widely applied.

**Acknowledgements**

This work was supported by the National Science Council (NSC), Taiwan, ROC, under contracts No. NSC 100-2511-S-006-014-MY3, NSC 101-2511-S-041-001-MY3, NSC 101-3113-P-006-023-, and NSC 100-2511-S-006-015-MY3.
References


The Effect of Intrapsychology Learning Before and After Interpsychology Activities with a Web-based Sharing Mechanism

Guo-Liang Hsu and Wu-Yuin Hwang*

Graduate Institute of Network Learning Technology, National Central University, Taiwan // Graduate Institute of Network Learning Technology, National Central University, Taiwan // hsukore@gmail.com // wyhwang@cc.nctu.edu.tw

*Corresponding author

(Submitted November 19, 2012; Revised March 8, 2013; Accepted April 20, 2013)

ABSTRACT
Learning is predominantly contributed through the approach, “first interpsychology, and then intrapsychology” called “intra- after inter-psychology.” This learning method was also defined in the “Zone of Proximal Development” from the assistance provided by others to the self, as proposed by Vygotsky. In this study, another reverse learning approach, “first intrapsychology, and then interpsychology” called “intra- before inter-psychology” was designed, so as to investigate what kind of learning sequence activities can bring about more effective learning experiences. Furthermore, the Web-based sharing annotation mechanism was employed to facilitate intrapsychology activities. Experimental results from a social course with 151 fifth-grade elementary school participants indicated that the intrapsychology activity plays an important role in students’ learning performance, regardless of whether it was structured before or after interpsychology activities. However, formulating learning concepts through textual annotations made in the intrapsychology activity, undertaken before the interpsychology activity, created a psychological tool for mediating interaction in interpsychology activities. In addition, the Web-based sharing mechanism could empower participants’ willingness to engage in the intrapsychology activity if the sharing mechanism was conducted before class. Constructing a structure for interpsychology activities has been explored for decades as an effective way to facilitate interaction among group members or to promote learning performance; however, this study further implies that conducting the learning strategy within the intrapsychology activities may be another approach worth further investigation; this is especially true for the strategy designed before interpsychology activities. Implications for practice and recommendations for future research are presented.

Keywords
Intrapsychology learning, Interpsychology learning, Web-based sharing mechanism, Annotation tool

Introduction
The central concept of Vygotsky’s theory is that individual cognition is based on the interaction among peers and their environment; the interaction is interceded by a mediator, such as tools, signs, and symbols. The mediator performs a vital role during the transformation from an inter-personal experience into an intra-personal form (Vygotsky, 1978). According to the socio-cultural approach proposed by Vygotsky (1978), learning is mainly an approach through interpsychology and intrapsychology. Vygotsky (1978) believed “every function in the child's cultural development appears twice; first, on the social level, and later, on the individual level -- first, between people (interpsychology), and then inside the child (intrapsychology).” Interpsychology refers to interaction with others (Berge, 1999), such as “Questions and Answers” between students or between the instructor and students in class. Intrapsychology refers to the internalization, or reflection of knowledge, such as summarizing or annotating learning materials through self-study before or after class.

The sequence of learning activities described briefly above is similar to the sequence the “Zone of Proximal Development (ZPD)” stages proposed by Vygotsky (1978). This sequence is the distance between where performance is assisted by more capable students and where performance is unassisted. Learning in the ZPD could be divided into several stages, which includes the assistance and learning regulation provided from others to the learner himself. Thereby, the sequence of learning activities within a ZPD model serves as the first social interaction with capable peers (a kind of interpsychology), then self-regulation by the learner himself (a kind of intrapsychology). This sequence of learning activities is “intra- after inter-psychology” activities. For example: first the teacher presents a lecture, then interaction between peers in class takes place, and then ‘self-study’ activities (called review) follow after class.
Many literatures showed the relevance and importance of intrapsychology or interpsychology modes on learning. An “intra- after inter-psychology” learning sequence attained better learning effects than modes that only used in-class activities. For example, modes that used intrapsychology after interpsychology activities resulted in better learning achievement, facilitated better understanding and retention of learning materials, and attained higher levels of student confidence and creativeness (Cooper & Valentine, 2001; Hwang et al, 2011; Hwang, Wang, & Mike, 2007; Ongun, Altas, & Demirag, 2011).

There is another common reversed sequence in classroom learning, “intra- before inter-psychology” activities. An example of this sequence is a self-study before class (called pre-reading), followed by a lecture and peer discussion activities in class. The “intra- before inter-psychology” learning modes obtained better learning achievement than modes that employed only in-class activities. Namely, the top achievements were: students’ learning achievement levels were improved and students’ attention was more focused in class on the part previously not fully understood during pre-reading activities (Chen, 2008; Chiu & Lee, 2009). Web-based technology with a sharing mechanism into “intra- before inter-psychology” activities was also used to gauge the effectiveness of pre-reading on learning achievement and peer collaboration (Cobos & Pifarré, 2008; Hwang & Hsu, 2011).

However, few studies concentrated on further comparison between the learning sequence mode with “intra- before inter-psychology” activities and the mode with “intra- after inter-psychology” activities on learning, particularly within school classrooms. Furthermore, the effectiveness and comparison of “intra- before inter-psychology” activities and “intra- after inter-psychology” activities with/without a Web-based sharing mechanism were seldom studied.

Using the concepts of “interpsychology and intrapsychology” as a theoretical framework, this study addresses the following research objectives through a Web-based annotation system: (1) the effects of different activity sequence modes with inter/intra-psychology on learning achievement were investigated; (2) activity sequence modes with a Web-based sharing mechanism were conducted along with their influence on learning achievement and the quantity of annotations; (3) the participants’ perceptions about activity sequence modes with inter-/intra-psychology, and the sharing mechanism were also explored.

The research questions that guided the study were as follows:
1. What were the effects of the sequence modes, intrapsychology learning conducted before or after inter-psychology activities, on learning?
2. What were the effects of the Web-based sharing annotation mechanism on learning?
3. What were the perceptions of the participants toward the intrapsychology learning before or after interpsychology activities and the Web-based sharing annotation mechanism?

**Literature**

**Influence of the sequence modes using an intrapsychology activity before and after an interpsychology activity on learning**

The learning sequence mode with “first pre-reading before class, then interaction in class” is an “intra- before inter-psychology” activity. Pre-reading has great influences on learning achievement. Preparation for learning before class helps students to be “learning-ready” (Paul, 1979). Ten-minute in-class preparation for the pre-test could stimulate students to work hard and get some knowledge of the learning materials before learning (Chan, 2005). Furthermore, the students who read a legal case before class displayed a greater understanding of the learning materials than the students who did not prepare for a pharmacy law course before class (Spies & Wilkin, 2004). Chen (2008) showed that pre-reading helped learners capture key points and incomprehensible concepts before class, so they could focus their attention on the parts of the lecture related to those key points and concepts previously not fully understood.

A few researchers applied technologies into pre-reading activities to enhance learning achievement. Viewing a video of lectures before class improved the high-school students’ basic image processing (Chiu & Lee, 2009). Engaging in pre-reading activities through a Web-based annotation tool (i.e., textual annotations are written before class using a Web-based annotation tool) could reveal students’ preparation before class as well as reflect their prior knowledge to help the instructor conduct appropriate learning activities (Hwang & Hsu, 2011). Additionally, Hwang and Hsu (2011) further indicated that the quantity of the textual annotations written before class positively correlated with
learning achievement. Furthermore, a pre-reading strategy could increase learning achievement compared with classes that did not engage in self-study before class. Sun and Huang (2005) conducted an experiment with an experimental group and a control group, employing traditional instruction with or without requesting pre-reading via Web-based learning materials. The results indicated that the pre-reading group received significantly higher scores than the group that did not self-study before class.

The learning sequence mode that featured “first interaction in class, then review by oneself after class” is an “intra-after inter-psychology” activity. Reviewing activities, such as doing homework, have a positive impact on students’ learning. Furthermore, such activities have immediate effects on the retention and understanding of the learning materials (Cooper & Valentine, 2001). Homework with problem-solving in a cyber socio-cultural environment gave students the opportunities to explore learning materials, which bolster confidence and creativity (Ongun et al., 2011). The research of Hwang, Wang, and Mike (2007) indicated that the quantity of textual annotations written after class in the high achievement group (top 27 percent of the experiment class) was significantly higher than the number of annotations written by the low achievement group (bottom 27 percent of experiment class). Meanwhile, the quantity of textual annotations written after class had a positive relationship between learning achievement. Therefore, it was implied that the more textual annotations participants created after class would affect their learning achievement whether he/she was in a low or high achievement group. Reviewing textual annotations written by original annotators has a significant influence on learning achievement. However, inexplicit annotations did not significantly influence learning achievement because textual annotations act as meaningful symbols for original annotators (Hwang et al., 2011).

Both intra- and inter-psychology learning sequence modes had positive learning effects for students. Previous research related to annotation tools seems to reveal that the quantity of textual annotations serves as a factor that correlates with participants’ learning performance when used to explore the effectiveness of Intra psychology before or after class. Detailed empirical studies are needed to further investigate the effectiveness and to compare the usefulness of different sequence modes on learning and the students’ attitudes toward the modes.

**Influence of the Web-based annotation sharing mechanism in an “intra- before inter-psychology” activity and an “intra- after inter-psychology” activity on learning**

Annotation refers to making extra information on documents. Annotation contains two different type forms: writing comments on a document is considered explicit annotation, and underlining or making highlights is considered inexplicit annotation (Marshall, 1997). Explicit annotations function as short notes of interpretation, reflections for a visible trace of annotators’ attention (Marshall, 1997). Therefore, explicit annotations could be a means for carrying out intrapsychology activities, either before or after interpsychology activities. Meanwhile, the previous findings also indicated that the more explicit annotations made by an annotator, the higher learning achievement they achieve because the explicit annotations create greater meaning for the annotator than the inexplicit ones in interpsychology activities (Howe, 1997; Hwang et al, 2007; Hwang et al., 2011).

Using a Web-based annotation sharing mechanism to check the learning effects of “intra- after inter-psychology” activities has been used by educational psychologists for decades to promote the following effects: to stimulate students’ motivation to engage in learning activities and to help them move forward toward learning goals (Hwang, et al., 2007); to achieve better learning performance by collaborating with peers via sharing and accessing their own ideas of the learning materials used (Hwang et al., 2007; Su, Yang, Hwang, & Zhang, 2010); to obtain the benefits of peer learning, such as offering more learning opportunities through conversation or dialogue via sharing annotations (Glover, Xu, & Hardaker, 2007); to learn a peer’s methods of how to accomplish a specific task via sharing and discussing their annotations with other peers (Cobos & Pifarré, 2008); and to support learner-center collaborative learning for adult and adolescent students via a shared document-based annotation tool (Nokelainen, Miettinen, Kurhila, Floréen, & Tirri, 2005).

Meanwhile, research of the effectiveness of a Web-based collaborative annotation system to attain better learning achievement, compared with the group using a discussion board system to read learning materials has been conducted (Su et al., 2010). However, the results by Hwang, Chen, Shadiev, and Li (2011) showed that using a sharing annotation mechanism to view classmates’ annotations marked on learning materials did not have significant positive effects on math learning achievement. This finding was due to the fact that reviewing peers’ textual annotations did not promote reviewers’ inspiration, as compared with viewing the textual annotations created by original annotators. Contrarily, reviewing peers’ textual annotations of homework solutions significantly influenced reviewers’ learning achievement.
Constructing the structure of interspsychology activities has been explored for decades as an effective way to facilitate interaction among group members, or to promote learning performance (Brwon et al., 1993; Collins, 2006; Stahl, Koschmann, & Suthers, 2006; Slavin, 1996; Zhang, Scardamalia, Reeve, & Messina, 2009). Expertise was mainly distributed through two activity forms of collaborative learning, jigsaw and reciprocal teaching (Brwon et al., 1993). The structures of instructional design models of jigsaw and reciprocal teaching were conducted to increase interdependence or interaction among members to facilitate rethinking and to promote the contribution of each member (Slavin, 1996). Meanwhile, social structure in collaborative learning also influences the contribution of members. For example, opportunistic collaboration based on emergent goals causes higher level contribution of each group member, as compared with the social structure with fixed groups (assign each division of the task to members and combine the contribution together at the end) or interacting groups (based on fixed groups and interaction between groups) (Zhang et al., 2009). The technology environment was applied as a tool to facilitate thinking and to shape thought in the process of interaction of disturbed expertise (Brwon et al., 1993). Computer Supported Intentional Learning Environments (CSILE) Project was the first system for networked collaboration learning (Bereiter & Scardamalia, 1993), and the upgraded version is Knowledge Forum (Scardamalia, 2004). A technology with a note-taking, note-viewing, and note-sharing mechanism was provided to support the interaction of disturbed expertise (Bereiter & Scardamalia, 1993; Scardamalia, 2004). Expertise is formulated during the process of collaborative problem-solving in which continuously sharing, thinking, and redefining notes inside CSILE/Knowledge Forum by contributing pieces of their work (Bereiter & Scardamalia, 1993).

In contrast with the Web-based sharing mechanism that was widely used in the learning sequence mode with “intra-after inter-psychology”, very little literature applied a Web-based sharing mechanism in the learning sequence mode with “intra- before inter-psychology” activities. A pre-reading sharing mechanism with Web-based annotation capabilities could stimulate and help students perform more useful pre-reading by reviewing others’ annotations, thereby enlarging the effectiveness of pre-reading as it relates to learning (Hwang & Hsu, 2011).

A Web-based sharing mechanism used in both learning sequence intra- and inter-psychology modes had positive learning effects for students. Meanwhile, previous research seems to reveal that conducting strategies with the support of a Web-based sharing mechanism may promote interaction between peers or increase the contribution of each member if conducted during interspsychology activities or intrapsychology learning after interspsychology activities. Detailed empirical studies are needed to investigate and compare the usefulness of a Web-based sharing mechanism on learning, and the interaction between peers as conducted in different sequence modes with intra-/inter-psychology activities.

Perceived usefulness and ease of use of system

Perceived usefulness and perceived ease of use were proposed by Davis (1986). The belief that using technology will increase and improve performance fulfilled the “perceived usefulness” category. The “perceived ease of use” category referred to the belief that using an information system will be free of effort (Davis, 1986). It was widely used to predict user attitudes toward information technology (Chang & Yang, 2010; Park, 2010). A person’s behavior toward an information system was determined by his attitude concerning perceived usefulness and perceived ease of use (Davis, 1986).

Therefore, in this study, the perceived usefulness and perceived ease of using a Web-based annotation system, called Virtual PEN (VPEN), with “intra- before inter-psychology” and “intra- after inter-psychology” activities were employed. Furthermore, the effects with/without Web-based sharing for learning in the different learning sequence modes were deeply investigated and the reasons behind findings were also analyzed.

Research method

Participants

Five classes with 151 fifth-grade elementary school students were involved in this study. As shown in Table 1, Class 1 and Class 2 were composed of pre-reading classes, where participants were asked to engage in self-study Web-based learning materials before class and to join in-class activities. Participants in Class 2, referred to as the Web-based sharing pre-reading class, were asked to self-study learning materials before class and then share their annotations with each other during pre-reading. Class 3 and Class 4 were review classes, where participants were
asked to self-study Web-based learning materials after class. Participants in Class 4, called the Web-based sharing review class, shared their annotations with each other during a review period and then self-studied learning materials and answered guiding questions. Class 5 was a traditional class, whose participants only joined in-class activities using a paper-based textbook and students were not asked to self-study before or after class. Therefore, Class 1 and Class 2 shared a similar learning sequence mode with “intra- before inter-psychology” activities, while Class 3 and Class 4 were organized by a similar mode with “intra- after inter-psychology” activities.

### Table 1. Groups, classes, and participants

<table>
<thead>
<tr>
<th>Group:</th>
<th>Group 1: intra- before inter-psychology group (Class 1 and Class 2)</th>
<th>Group 2: intra- after inter-psychology group (Class 3 and Class 4)</th>
<th>Group 3: Traditional group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class:</td>
<td>Class 1: Self-study before class</td>
<td>Class 2: Self-study with sharing mechanism before class</td>
<td>Class 3: Class 4:</td>
</tr>
<tr>
<td></td>
<td>Class 3: Self-study after class</td>
<td>Class 4: Self-study with sharing mechanism after class</td>
<td>Class 5: Traditional class with only in-class activities</td>
</tr>
<tr>
<td>Participants:</td>
<td>30 students</td>
<td>30 students</td>
<td>30 students</td>
</tr>
</tbody>
</table>

### Experimental procedures

An eight-week quasi-experiment, with each week devoting two hours to research, was conducted to investigate the effects of the treatment on learning achievement and the participants’ perception toward the approach. The experimental procedures are described as follows (and illustrated as Figure 1):

1. **Training activities:** According to Table 1, training courses for the use of VPEN with or without a Web-based sharing annotation mechanism, pre-reading, and review activities were provided.
2. **Pre-test:** Before the treatment, a pre-test was held for all classes to measure the students’ prior knowledge.
3. **Treatment:** Different learning sequence modes with inter-/intra-psychology activities and a Web-based sharing mechanism.
1. Pre-reading activities (self-study before class):

   Participants in Class 1 and Class 2 were asked for making self-study before class two weeks ago. The Web-based sharing annotation mechanism was provided for those in Class 2 since the self-study activities were performed for one week. The task for self-study activities included reading and making annotations on the online textbook materials, and completing the guiding questions designed by the instructor.

2. In-class activities:

   Classmates were divided into several groups; each one had four to five members. Each group had both male and female students with different ability levels depending on their prior knowledge, based on pre-test scores. Each class had the same structure: for the first 5 minutes was devoted to the instructor’s guiding time, the middle 25 minutes was group discussion time for a problem-based task designed by the instructor, and the final 10 minutes were dedicated to individual activities, in which students wrote down their own solutions for the tasks. The classroom for all classes in this study was the computer classroom, which was accessible for content annotation before class or to search the internet.

3. Review activities (self-study after class):

   The review activities were organized after class for two weeks in Class 3 and Class 4. The Web-based sharing mechanism was provided for participants in Class 4 for one week after class. The task to complete for review activities is the same for pre-reading activities.

4. Post-test: After two weeks of the treatment, a post-test was held for all classes to measure the students’ learning achievements.

5. Questionnaire and interview: At the end of the experiment, open-ended questionnaires and interviews were provided to identify the participants’ perceptions toward inter-/intra-psychology activities and the sharing mechanism.

   Data collection and analysis: Scores from the pre-test and post-test, the quantity of annotations that were calculated by the concept-based coding, and the open-ended questionnaire and interview were collected. Methods of analyzing the data included Analysis of Covariance (ANCOVA) (used to examine the effects of different learning sequence modes), the Pearson correlation (used to examine the correlation between learning achievement and the quantity of explicit annotation), etc.

Learning materials, questionnaire, and interview

Five classes had the same learning materials, which included a social course textbook in a fifth-grade elementary school. According to the concepts, perceived usefulness, and perceived ease of use proposed by Davis (1986), the open-ended questions designed by the researchers were the follows.

1. What do you think is the usefulness of the Web-based annotation system on learning?
2. What do you think is the ease of use of the Web-based annotation system?
3. What do you think is the usefulness of pre-reading/review on learning?
4. What do you need while you study before or after class?
5. What do you think is the usefulness of the Web-based sharing annotation mechanism on pre-reading/review and learning?

   Interviews were only applied when the questionnaire responders wrote their position toward the proposed activities, but the reasons were not clearly presented. For example, “I think self-study before class was useless for my learning. . . .”

Results and analysis

Effects of different activity sequence modes and the Web-based sharing mechanism on learning achievement

ANCOVA was adopted for examining the possible effects of different activity sequences with inter-psychology and intra-psychology and the effects of the Web-based sharing mechanism on learning achievement. First, the ANCOVA
results are presented in Table 2. These results show that Group 1, whose participants engaged in an “intra- before inter-psychology” activity sequence, significantly outperformed Group 2 (with an “intra- after inter-psychology” activity sequence) and the traditional group (Group 3); and Group 2 significantly outperformed Group 3. The results further reveal that an “intra- before inter-psychology” sequence achieves more positive effects on participants’ learning than the activities with reverse order. Self-study would be more helpful for participants if it is organized before the interaction among peers or the instructor. Additionally, the traditional group, whose students only engaged in in-class activities, demonstrated the least effect on learning achievement as compared with Group 1 and Group 2.

Table 2. Effects of Sequence Mode with interpsychology and intrapsychology Activities on Learning Achievement

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Mean</th>
<th>SD</th>
<th>Adjusted mean</th>
<th>F(2,147)</th>
<th>p</th>
<th>G1&gt;G2*</th>
<th>G1&gt;G3*</th>
<th>G2&gt;G3*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intra before Inter psychology, Group 1</td>
<td>60</td>
<td>83.03</td>
<td>14.805</td>
<td>83.736</td>
<td>21.377</td>
<td>.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intra after Inter psychology, Group 2</td>
<td>61</td>
<td>77.10</td>
<td>15.687</td>
<td>76.690</td>
<td></td>
<td></td>
<td>G1&gt;G3*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inter psychology only, Group 3</td>
<td>30</td>
<td>62.30</td>
<td>16.795</td>
<td>61.726</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p < 0.05.

Second, the analysis of ANCOVA is further applied to investigate the effect of different modes with and without a Web-based sharing mechanism. The findings, as shown in Table 3, reveal the following:

1. Self-study with a sharing mechanism before class (Class 2) did significantly better than all classes. No significant difference was found among others (Class 1, Class 3, and Class 4). Thereby, the Web-based mechanism applied in self-study before class is helpful for participants’ learning as compared with the other sequences with or without a sharing mechanism.

2. The results also reveal that in the review classes (Class 3 and Class 4) the Web-based sharing mechanism did not significantly impact participants’ learning because the learning achievement between classes with and without this mechanism was not significantly different.

3. Furthermore, Class 1, Class 2, Class 3 and Class 4 did significantly better than Class 5, which reveals that only in-class learning in Class 5 -- without applying the mechanism of self-study before or after class -- performed the poorest out of all classes.

Table 3. Effects of different sequence modes with/without web-based sharing mechanism on learning achievement

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Mean</th>
<th>SD</th>
<th>Adjusted mean</th>
<th>F(4,145)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-study before class and in-class activities, Class 1</td>
<td>30</td>
<td>78.60</td>
<td>15.420</td>
<td>79.652</td>
<td>12.976</td>
<td>.000</td>
</tr>
<tr>
<td>Self-study with sharing mechanism before class and in-class activities, Class 2</td>
<td>30</td>
<td>87.47</td>
<td>12.942</td>
<td>87.794</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-study after class and in-class activities, Class 3</td>
<td>30</td>
<td>74.00</td>
<td>15.186</td>
<td>73.414</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-study with sharing mechanism after class and in-class activities, Class 4</td>
<td>31</td>
<td>80.10</td>
<td>15.823</td>
<td>79.874</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Only in-class activities, Class 5</td>
<td>30</td>
<td>62.30</td>
<td>16.795</td>
<td>61.736</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p < 0.05.

The above findings probably imply the following:

An intrapsychology activity plays an important role in students’ learning performance, whether it was arranged before or after interpsychology activities, as compared with the performance without intrapsychology activity. If the Web-based sharing mechanism was applied during the intrapsychology activity, the “intra- before inter-psychology” activity might obtain better learning performance than the “intra- after inter-psychology” activity. The Web-based sharing mechanism did not facilitate the majority of the effects on learning achievement, as it was applied during the intrapsychology activity, which was after the interpsychology activity. However, it might expand the significant difference of learning performance during the reverse sequence.
Correlation between learning achievement and the quantity of annotations in different sequence modes

According to the findings presented in Table 2 and Table 3, learning sequence modes with intrapsychology organized before or after an interpsychology activity would significantly effect participants’ learning achievement compared with the activity without intrapsychology activity. Therefore, further investigation of the correlation between learning achievement and the effort participants put into an intrapsychology activity (the quantity of explicit and implicit annotations in this study) is necessary.

The Pearson correlation between learning achievement and the quantity of explicit annotations showed that significant correlation existed ($r = .198$, $p = .030 < .05$, $N = 121$), while the inexplicit annotation did not ($r = .05$, $p = .583 > .05$, $N = 121$). This finding reveals that high values of participants’ learning achievement tended to be positively associated with the large values of the explicit annotation quantity in the intrapsychology activities. If students study enough through explicit annotations during intrapsychology learning, they may gain higher learning achievement.

The usefulness of a Web-based sharing mechanism, which was applied in intrapsychology activities, on the quantity of explicit annotations was further investigated. The Post Hoc comparison of ANOVA shows that with the exception of the significant difference between those engaged in self-study with a sharing mechanism before class (Class 2) and those who engaged in self-study before class (Class 1) ($Mean Difference = 9.27$, $Standard Error = 2.877$, $p = .019 < .05$), no significant difference was found among other classes. The quantity of explicit annotation in each class was ranked from high to low: the Web-based sharing pre-reading class, the review class, the Web-based sharing review class, and the pre-reading class (Class 2: $Mean = 24.27$, $SD = 15.2$, $N = 30$; Class 3: $Mean = 21.07$, $SD = 10.907$, $N = 30$; Class 4: $Mean = 17.48$, $SD = 7.953$, $N = 31$; Class 1: $Mean = 15$, $SD = 9.24$, $N = 30$).

The findings reveal that the more students were engaged in intrapsychology learning through the creation of more explicit annotations, the higher learning achievement attained. The quantity of explicit annotations in self-study before class was ranked the least, which reflects that self-study before class without a sharing mechanism may be not easy for some participants. The Web-based sharing mechanism could help participants make more explicit annotations compared with the self-study before class without a sharing mechanism. No significant differences between the review class and the Web-based sharing review class reveal that during the review period, the Web-based sharing mechanism did not assist participants in making more explicit annotations. Based on the open-ended questionnaire and interview, participants felt that reading annotations made by classmates was unnecessary due to the perception that some concepts about learning materials were understood after hearing the lecture or after discussion in class.

After investigating the influence of intra-/inter-psychology activities on learning achievement and the correlation between learning achievement and the quantity of annotations written during intra-psychology activities, it is necessary to further explore the usefulness of a Web-based sharing annotation mechanism and explicit annotations written in different intrapsychology activities (before or after interpsychology activities) on learning.

Analysis of the open-ended questionnaires and interviews

The questionnaire was given to four classes with 121 participants, and 101 completed questionnaires were received (an 84 percent return rate). A total of 40 participants were interviewed to verify their attitudes toward the treatments.

Integrated analysis of perceived usefulness of intrapsychology annotation for learning performance

The usefulness of pre-reading was perceived as follows:
1. The process of reading and creating explicit annotations during pre-reading activities facilitated participants to acquire some concepts, to think, or to find some questions before class. Their learning concepts formulated before class, which may be different from or conflict with others, led to further discussion in class, which, in turn, resulted in better learning achievement. For example, some participants (who were in a self-study class without a sharing mechanism before class) responded to the usefulness of explicit annotation on learning:
“I annotated some learning questions, abstracted in each paragraph before class. The learning questions or abstract (in textual annotations) written by me or my classmates before class could be a topic or guideline to help us effectively discuss in class. I think effective discussions help me grasp the key points and leads to an increase in my score on the final exam.”

“I wrote down some reflections after I read the lesson on VPEN before class. I thought I already knew some key points, but after discussion in class I found part of them were not correct, thanks to my friends’ suggestions. It helps me a lot.”

Other participants responded to the process of creating explicit annotations led to more active interaction in class and resulted in better learning achievement. (The following perceptions come from some participants whose scores were higher through the use of self-study with a sharing mechanism before class.)

“After I read the annotations written by my classmates, I found they were different from mine sometimes. It made me think again and again. Carefully rethinking about what is different between them before class helped me be able to describe my opinions and to more actively discuss with my classmates in class. The active discussions help me to clarify some misunderstandings about the lives of aborigines. Therefore, I get better scores.”

Another participant had a similar response.

“After I read the textual annotations written by some classmates before class, they considered aborigines in our county who like to dance and drink. My classmate’s opinion about how the aborigines use dance and drink to fight with enemy interested me. What kind of crazy guy wrote this comment? It is impossible to use them against an enemy...After a long discussion, we finally came to the conclusion that aborigines like to dance, drink, and draw on their faces because they like to make themselves appear brave to fight with enemies. They also use these things to celebrate victory. A colorful drawing, like a bridge, brings them to heaven and forces the enemy to hell. I think it’s their culture and custom. We should respect it. The discussion also let me better understand the colorful lives and customs of aborigines, and it helped us get better scores on the final exam.” (This lesson pertained to colonial expansion and aborigines)

However, another culture and custom is not easy experience, particularly for the new generation that did not have the context to experience. Their diverse learning concepts formulated in the process of writing explicit annotations before class further contributed to formulate the abstract concept of culture and custom in the class discussion. For example, some classmates used some annotations about aborigines’ objects written before class (a ship with totems, a symbol of snakes, clothes with colorful symbols) to cooperatively compose a story.

“The totems on the ship look like a star or sun. These symbols take us to the ocean and safely return us back. The totems are like eyes for the ship to direct the way home. The totems on the ship look like fishing net to catch fish. The totems are like eyes for the ship in search of fish. Snakes are monsters in the ocean. Beautiful totems on the clothes resemble a net to prevent the attack of a water snack. The totem net looks like the skin of the snack and it drives us home safely and directly.... I thought the story was an example of how to reply to the open questions related to the life of aborigines in the final exams.”

The story the students composed likely did not completely match the original spirit of the culture. But it might imply that the aborigines’ objects came alive again in their story, and social meanings were reinterpreted, reshaped by them, but the spirit is the same, telling the new generation to have respect for their ancestors and nature.

2. Finding some key points before class and interacting with classmates and the teacher in class deeply impressed participants (the first time involved writing explicit annotations in pre-reading, and interaction in class was the second) and produced better learning achievement. For example, some participants responded:

“Reading materials, writing some main points before class, and then listening and discussing these points in class makes me remember the materials.”
3. The more preparation the participants had before class, the more confidence participants had during the activities designed in class. For example:

“I have confidence to correctly answer the questions asked by peers or the teacher if I have annotated enough comments after reading the materials before class.”

“If I read materials before class, it helped me more easily realize what my classmates discussed in class. It is good for me and I can express my opinions about the questions. It is exciting for me.”

“I am nervous to express my opinions in public because I have no confidence in my opinions. But after I read the annotations written by my friends, I am not nervous anymore because some of their opinions are similar to mine and I felt my opinions are better than theirs.”

4. The limitation of class time causes participants to have limited opportunities to explain their opinions that conflict with peers. Meanwhile, only offering guiding questions and arranging pre-reading learning schedules by the teacher were not useful or they did not satisfy some participants. For example, some participants did not feel at ease during self-study before class. Another participant, with high prior knowledge, felt bored and frustrated that the other students’ learning progress was not known. Also, those with low prior knowledge described that they needed more help from others to learning the concepts (this analysis was from the pre-reading class without using a Web-based sharing annotation mechanism).

“Where can I find resources or solutions to solve the guiding questions? I need some direction on how to handle this.”

“It is boring. I have already written my solutions or comments. But I want to know what my classmates wrote and whether they study hard.”

The responses provided in the open-ended questionnaires and interviews probably reveal the participants’ perceived usefulness of review on learning in the following ways:

1. Reviewing learning materials and annotations helps participants to remember what they learned in class. The lecture from the teacher or the opinions of peers helps participants to quickly determine main ideas. Reviewing the annotations made in class helps the students memorize the main ideas of the learning materials.

2. The review activities offered participants opportunities to confirm what they learned. The process of completing some guiding questions designed by the teacher gave the participants the opportunity to ensure that they accurately realized the concepts and confirm what they learned. Some participants responded that they still did not understand the concepts, or they did not grasp the main ideas despite the opportunity to review the learning materials. They need more guidance from peers or the teacher.

3. The learning schedule designed by the teacher for one class may be different for a subsequent class. Some participants feel embarrassed to ask for help in one class due to the timing of finding the learning questions; they may be unable to catch up with the learning schedule in class even though the teacher offered some time to ask for help in class.

**Integrated analysis of a Web-based sharing annotation mechanism on facilitating intra-/inter-psychology learning**

The usefulness of a sharing mechanism applied in intra-psychology learning before inter-psychology activities:

1. The sharing mechanism facilitates the engagement of participants’ interaction in class.
   
   (1) Reading the explicit annotations shared by classmates not only helped participants grasp more main ideas, but it also helped them identify some learning problems that needed further verification. That made them pay attention to the teacher’s lectures or peers’ opinions in class.

   “It helps me more easily and quickly joins the discussion with my classmates. Before the discussion, I read my classmates’ annotations and they remind me of something I did not pay attention to before and they also let me know there are some topics I should further discuss and make clear in class.”
(2) Reading the explicit annotations written by classmates helped participants to grasp the structure of materials before class, which helped them to catch up with the pace during interaction activities. For example:

“Reading the annotations made by my classmates helped me know the structure of the lessons before class, which helped me to actually know what we are discussing right now, and what would be discussed next in class discussion.

2. The sharing annotations mechanism motivated participants to be more engaged in intrapsychology activities.

(1) The mechanism provided opportunities for improving the contents of annotations: Peers’ explicit annotations might contain some related learning resources, such as references of websites. These annotations gave some participants a new direction to improve their answers to the pre-reading guiding questions. For some low prior knowledge participants, classmates’ annotations served as annotation practice models. Therefore, sharing, and viewing an annotation mechanism supported some participants to continue to study learning material before class.

“My classmates would annotate some questions to remind or guide themselves when something is unclear. The annotations also provided some direction on what I need to improve. Particularly, the references they annotated guided me to a website that helped me to better understand how to answer the guiding questions.”

Another participant responded that reading others’ annotations inspired him to find more useful information before class.

“I saw textual annotations written by my friends before class. A classmate said she interviewed her grandfather to gain a better understanding of the culture of aborigines. Another friend expressed that he watched a television program to realize the customs of aborigines. I went to the museum of aborigines to access their dedicated collections. Watching the concrete collections help me experience the daily life of aborigines. I heavily depend on the internet to search for information, but after reading classmates’ annotations, I found it was not always the case, culture, and custom of aborigines’ lives. After discussion in class with my classmates, we thought the best way to experience the culture and custom was to access their life stories.”

(2) The mechanism enabled a way to easily monitor the learning process of classmates: Participants could easily monitor how many annotations classmates had written through the “Switch Users” menu of the VPEN system, as shown in Figure 2. For some high prior knowledge participants, reading peers’ annotations gave them insight to their competitors’ learning progress, which motivated them dedicate more time to study the learning materials before class. For example, some participants responded:

“I could use it to watch how much effort my competitor put into studying and what they did. This made me study harder to get a better learning achievement. It is inconvenient to read all competitors’ annotations written in paper-based textbooks before class.”

(3) The mechanism helped participants reflect on annotations. Participants thought twice before making annotations and they reflected on what they annotated. This reflection motivated them to be eager to join the discussion among classmates in class.

“Sometimes I found annotations made by my classmates were not the same as mine and sometimes I am not sure mine is correct. I would rethink the details. I know what part of my annotation may have errors. I am eager to join the discussion in class and tell others about my opinions. I thought checking and reflecting made me get a better score.”
The usefulness of sharing mechanism applied in intrapsychology learning after interpsychology activities:

Few participants were satisfied with the usefulness of the Web-based sharing mechanism in the intrapsychology activities after class. This was because after the teacher’s lecture and open discussion with peers in class, some participants perceived that they obtained some key ideas. As a result, it was not necessary to read others' annotations. Meanwhile, the participants thought the main ideas found by peers were not better than the explanations provided by the instructor; therefore, they paid attention to their own annotations made in class instead of reading peers’ annotations.

Integrated discussion of relationship among aspects

According to the results above, the relationship among aspects (a Web-based sharing annotation mechanism, intra-/inter-psychology learning, perceptions among participants, quantity of explicit annotations and learning achievement) is presented in Figure 3.

First, both the engagement in intrapsychology activities through the creation of explicit annotations and the sequence modes with intra-/inter-psychology activities would likely affect learning achievement:

- In this study, the quantity of annotations was referred to as the effort of participants engaged in an intrapsychology activity. The positive significant correlation between the quantity of explicit annotations and learning achievement probably implies that if participants were more engaged in intrapsychology activities through the creation of explicit annotations, better learning achievement was likely to follow. This may partly explain why the Web-based sharing pre-reading class had the highest learning achievement -- its quantity of explicit annotations ranked first among all classes.

- However, an interesting phenomenon was found in that no significant difference in the quantity of explicit annotations was found between different sequence intrapsychology activities (the Web-based sharing pre-reading class and other review classes), however, the significant difference of learning achievement occurred between them. It perhaps also implied explicit annotations written during intrapsychology learning before interpsychology activities may more effectively increase learning achievement than those written during intrapsychology learning after interpsychology activities. The more engagement in intrapsychology activities before interpsychology, the higher the learning achievement attained.
The survey of perceptions reveals that explicit annotations and learning concepts formulated before class may act as a mediator to facilitate the rethinking of learning concepts and to prompt interaction among participants in class, which would likely prompt increased learning achievement. However, intrapsychology learning after interpsychology activities helps participants to master or adjust learning, but it does not effectively inspire learners to interact among members in class. Therefore, it seems that two factors (one is the learning concepts formulated in the intrapsychology learning; another is the different sequence modes of intra- and interpsychology learning) might influence the interaction between peers in interpsychology activities. The analysis of perception might explain why an “intra- before inter-psychology” activity sequence significantly outperforms “intra- after inter-psychology” activity sequence and an “intra- after inter-psychology” activity sequence significantly outperforms the activity with “inter-psychology only,” as shown in Table 2.

Second, the factor (the Web-based sharing mechanism) facilitated the engagement in intrapsychology activities conducted before interpsychology activities and promoted discussion during interpsychology activities:

- The Web-based sharing mechanism applied before class could be a factor that affected the positive attitude of participants towards intrapsychology learning. For example, participants, who scored higher in the pre-reading class without using the Web-based sharing annotation mechanism, felt bored. On the contrary, some high prior knowledge participants, who were from the pre-reading class that used the Web-based sharing annotation mechanism, were interested in monitoring the learning process of competitors and they were incited to be fully engaged in creating more annotations. The perceived positive attitude could explain why the quantity of annotations in Class 2 (self-study with sharing mechanism before class) was significantly greater than those who self-studied on their own without the sharing mechanism before class.

- This sharing mechanism, however, would not effectively inspire participants to be more engaged in intra-/interpsychology learning if it was conducted after class. For example, participants felt embarrassed to ask for help in the next class or they felt they should have been able to get the main points after interaction in class. Thus, it was not necessary to read classmates’ annotations via the Web-based sharing mechanism. This survey of perception explains the results -- why no significant difference in the quantity of explicit annotations between the self-study after class with/without Web-based sharing mechanism (Class 3 and Class 4) existed.

- Furthermore, the survey of participants’ perceptions reveals that the explicit annotations written before class would be suitable materials for further discussion in class. The Web-based sharing mechanism conducted before class likely prompted interaction in class. For example, reading explicit annotations shared by classmates likely caused concept conflicts before class, which led to them to rethink their annotation and to be eager to discuss opinions in class.

- Therefore, it seems that with the exception of the two factors (learning concepts formulated in the intrapsychology learning and the different sequence modes), which were discussed above, a Web-based sharing
mechanism would be another factor that might influence the interaction between peers in interpsychology activities. This analysis explains why the “intra- before inter-psychology” activity sequence with a Web-based sharing mechanism outperformed the most, which was probably because this sequence mode included these three factors to engage in formulating learning concepts, stimulating interaction between classmates, thereby, resulting in the best learning achievement.

Conclusion and implications

This study is aimed at exploring the effects of intrapsychology activities arranged before and after interpsychology activities. Meanwhile, the effects and the participants’ perceptions of a Web-based sharing mechanism, which was used in intrapsychology activities, were also explored on learning. The major findings and conclusions are as follows.

First, learning concepts formulated through intra-psychology learning before inter-psychology activities acted as a mediator to facilitate interaction among members, resulting in better learning performance:

Previous researchers indicated that learning activities combined with intrapsychology activities arranged before or after interpsychology activities attain better learning performance than the activity sequence mode that use only interpsychology activities (Chiu & Lee, 2009; Cooper & Valentine, 2001; Hwang & Hsu, 2011; Hwang et al., 2011; Hwang et al., 2007; Ongun et al., 2011; Sun & Huang, 2005). The finding in this study further indicated that the intrapsychology activities arranged before class resulted in significantly better learning effects than the activities arranged after class. According to Vygotsky’s theory, mediation is the key to facilitating human learning and the psychological process (Kozulin, 2003). If a learner masters the important concepts (mediators) written by the instructor before teaching, they might gradually transform the cultural of the social world (Hall, 2007). In this study, the concepts in the guided question and the pre-reading activities before class acted as a mediator for personal mental activity and to promote the willingness to interact with members in a community. If participants engage in learning preparation for the key instructional concepts through explicit annotations before class, they might formulate learning concepts. However, such preparation did not guarantee that all learning concepts constructed by learners before class were effective. In contrast to that, the concepts from participants were diverse and some learning problems were perceived. The opportunity for textual annotations and concept formulation before class could be regarded as one of the symbolic tools of mediation in mental activities and to mediate the interaction between individuals and the group members. Therefore, those who engage in learning preparation before class might potentially clarifies learning concepts and get better learning effects than those who do not perform pre-reading activities.

Furthermore, in this study the finding -- interaction between peers in Inter psychology activities potentially promoted learning performance -- was in line with the previous studies related with collaborative learning of distributed expertise (Bereiter & Scardamalia, 1993; Collins, 2006; Stahl et al., 2006; Slavin, 1996). Constructing the structure of interpsychology activities has been explored for decades as an effective way to facilitate interaction among group members, or to promote learning performance (Brown et al., 1993; Collins, 2006; Stahl et al., 2006; Slavin, 1996; Zhang et al., 2009). This study further indicated that learning concepts formulated within intrapsychology learning before interpsychology activities, using a Web-based annotation sharing mechanism, also be another effective way to facilitate the engagement of interaction in a classroom or it may potentially increase learning performance.

Second, intrapsychology learning after inter-psychology activities helps one to master or adjust learning, but it does not effectively inspire learners to interact among members:

The intrapsychology activities promoted the participants to master their learning, to confirm what have them learned, and to adjust their learning according to its confirmation. This result can be interpreted, according to the ZPD theory proposed by Vygotsky (1978), as learning regulated by others and the self also facilitates the level of capacity that might be developed. Therefore, in this study the intrapsychology activity after interpsychology activities achieves better learning performance than activities, which only include interpsychology in class, which is consistent with previous studies (Hwang et al., 2007; Su et al., 2010). However, the intrapsychology activity after interpsychology activities likely does not effectively inspire participants to join discussion activities in class because a lack of formulating learning concepts before class leads participants to spend too much time discussing what should be
prepared before class. Participants might feel embarrassed to ask for help in the next class with a different learning schedule.

Third, the sequence mode with a Web-based sharing mechanism conducted before class potentially produces a more effective influence on learning achievement:

The finding reveals that the engagement in intra psychology learning positively correlated with learning achievement. No significant difference in the engagement occurred between these classes (intrapsychology activities before class with Web-based sharing mechanism and two classes with a self-study activity after class); however, the learning achievement between them was significant difference. It reveals that the engagement in intrapsychology before class potentially produces a more effective influence on learning achievement as compared with the engagement in intrapsychology conducted after class. Although the findings in this study indicated that engaging in intrapsychology learning before class help to formulate learning concepts, which the act as a mediator to facilitate the interaction among members, the survey of perception from the pre-reading class without a Web-based sharing mechanism revealed that self-study before class (intrapsychology learning before interpsychology activities) was not easily satisfied by all participants. The further analysis indicated that a Web-based sharing mechanism was more helpful for engaging participants in the process of formulating learning concepts (stimulated rethinking before class and facilitated further interaction in class) while it was conducted in intrapsychology activities before class than that used in intrapsychology activities after class or the pre-reading without a Web-based sharing mechanism. This analysis of the role of a Web-based sharing mechanism not only explains the phenomenon, that a Web-based sharing annotation mechanism enlarged the effectiveness of pre-reading on learning as compared with pre-reading without conducting a Web-based sharing mechanism (Hwang & Hsu, 2011), but it also enhances the findings in this study: the “intra-before inter-psychology” activity sequence mode outperformed the activity sequence mode with “intra- after inter-psychology.

Based on the results and discussion above, the following implications were provided:

Implication of designing interaction activities in classroom:

Previous studies have demonstrated that forming fixed groups, interacting groups or flexibly combining groups based on emergent goals would facilitate interaction during discussion (Zhang et al., 2009). Meanwhile, organizing sequence steps to require all members involved would be an effective way to promote the interaction and contribution of every member during discussion, such as the sequence steps in jigsaw and reciprocal teaching (Slavin, 1996). However, this study further implies that conducting the learning strategy within the intrapsychology activities may be another method worth further investigation, especially the strategy designed before interpsychology activities. Therefore, it was recommended that conducting an intrapsychology activity before class to formulate learning concepts (act as a psychology mediator) would be an effective strategy for designing interactive classroom activities and to potentially increase learning performance.

Implication to compose annotations (acting as a psychological mediator in interpsychology activities):

This research indicated that annotations written in intrapsychology learning would serve as a psychological mediator to rethink and facilitate interaction in class. Additionally, it was demonstrated that explicit annotation has more positive effects on learning achievement than inexplicit annotation. Therefore, it can be inferred that a teacher might encourage students to put more emphasis on composing explicit annotations rather than inexplicit annotations as they engage in the process of formulating learning concepts during intrapsychology activities.

Implication for motivating students engaging in the process of formulating learning concepts before class:

The VPEN annotation system in this study mainly supported the intra- and inter-psychology activities. The production of explicit annotations was mainly influenced by the Web-based sharing annotations mechanism of VPEN, or the activities sequence modes in this study. Therefore, it implied that designing some learning strategies that use some sharing technologies before class would be an effective way to increase the engagement of students in the process of formulating learning concepts during intrapsychology activities. For example, the “Switch Users” menu item, as depicted in Figure 2, could be used to monitor the learning process of classmates during intrapsychology activities, which motivated students to be engaged in making more explicit annotations during intrapsychology learning.
This study has shown the importance of psychology mediators formulated before or after class in the relationship among the aspects (i.e., a Web-based sharing annotation mechanism, the activity sequence modes with intra-/inter-psychology, and learning achievement). In this study, it was indicated that a Web-based sharing mechanism could be an effective mechanism to facilitate engagement in intrapsychology learning, but a few classmates need more mechanisms to help them regulate their learning in the process of formulating learning concepts during intrapsychology activities. Therefore, an important direction for future research is to match more advanced mechanisms, such as learning monitors, to the needs and preferences of each individual to help students and instructors efficiently evaluate, monitor, and regulate learning in the process of formulating learning concepts during intrapsychology activities before or after class. Furthermore, the influence of a learning monitor strategy, with different sequence modes (intra-/inter-psychology activities) on students’ contribution and interaction in a classroom and subsequent learning achievement, warrants further investigation.

Acknowledgements

This study is partially supported by the National Science Council of the Republic of China, Taiwan under the contract numbers NSC 98-2511-S-008-005-MY3, NSC 101-2511-S-008-012-MY3 and NSC 101-2511-S-008-013-MY3.

References


The Effect of Instructional Techniques on Critical Thinking and Critical Thinking Dispositions in Online Discussion

Filiz Kalelioğlu* and Yasemin Gülbahar
Baskent University, Faculty of Education Baglica Campus Ankara, Turkey // Ankara University, Department of Informatics Gumusdere Campus Ankara, Turkey // filizk@baskent.edu.tr // gulbahar@ankara.edu.tr

*Corresponding author

(Submitted November 21, 2012; Revised March 11, 2013; Accepted May 3, 2013)

ABSTRACT

The aim of this research study was to explore the effect of instructional techniques on critical thinking and critical thinking dispositions in online discussion, based on triangulation design. Six Thinking Hats, Brainstorming, Role Playing, Socratic Seminar, and Anyone Here an Expert, were selected as an instructional techniques for online discussion. In the quantitative part, according to the results of ANOVA, except Socratic Seminar, there is no difference between groups in terms of scores of pre-tests and post-tests of critical thinking dispositions. In the qualitative part, according to the results of the analysis of critical thinking in online discussion, the Mixed Techniques group performed as having the best ability of critical thinking, the Anyone Here an Expert group was second and the Brainstorming group was third in terms of performing critical thinking ability in online discussion.

Keywords
Instructional techniques, Critical thinking, Critical thinking dispositions, Online discussion

Introduction

Higher order thinking skills like critical thinking, creative thinking and problem solving are considered necessary skills for 21st century individuals. On the other hand, technology competencies like using the Internet and its services effectively and learning in online environments are also skills required for the new generation. Hence, it is necessary and important to examine these dimensions from different points of view in order to develop ideas on the ways to best equip individuals with these skills and to make them more easily cope with emerging technologies and situations.

Although online teaching, or nowadays more commonly called e-Learning, is not a new phenomenon for the education world, few studies deal with both the implementation of instructional strategies and techniques within these virtual environments and the consequences of these various implementations in terms of enhancing critical thinking skills. We have to discover the best instructional techniques for students learning within online environments, as we do for face-to-face learning environments. As educators, we should consider the best practices which enhance students’ thinking skills, academic achievement, retention and other important dimensions of student learning. Owing to these facts, this research study was carried out to explore whether or not selected instructional techniques can enhance students’ critical thinking skills and critical thinking dispositions. Hence, six instructional techniques were used to conduct discussions in online environments in order to explore the critical thinking skills and critical thinking dispositions of students.

Critical thinking and critical thinking dispositions

Teaching students how to think critically is an essential issue in educational settings (Facione, 2007; Şendağ & Odabaşı, 2009) since critical thinking (CT) is very important to participate effectively in a democratic society with a set of skills in terms of workplace decision making, leadership, clinical judgment that affects directly professional success. As synthesised by Onions (2009) “Critical thinking is a way of thinking, and a set of skills, that encourages an informed, aware, systemic, considered and logical approach to deciding what to believe or do. Critical thinking leads to arguments and conclusions that are valid, substantiated and resistant to criticism” (p. 2). As also stated by MacKnight (2000), teaching CT by using online discussion is an essential approach in terms of enhancement of teaching and learning in electronic forums. Students should enhance their critical thinking abilities in order to cope
with the information explosion and other rapid technological changes that we are faced with in recent years and also for the upcoming years.

From a different point of view, Halpern (1999) underlined the importance of addressing student dispositions in terms of critical thinking instruction and stated that “Critical thinking is more than the successful use of the right skill in an appropriate context. It is also an attitude or disposition to recognize when a skill is needed and the willingness to exert the mental effort needed to apply it” (p. 72). Thus it is crucial to explore innovative ways to make our students value both good thinking and the effort that is necessary to use their skills. Yang and Chou (2008) explored the relationship between critical thinking skills (CTS) and critical thinking dispositions (CTD) and investigated the effectiveness of instructional strategies in improving students’ CTS and CTD. Although they found a positive relationship between CTS and CTD, they concluded that only the students with high CTS and medium CTD possessed a significant correlation. The researchers also concluded that in terms of effectiveness of the instructional strategies, both CTS and CTD need to be and can be taught and cultivated. Han and Brown (2013) conducted a study which is based on an intervention designed to improve early childhood teacher candidates’ critical thinking skills. Their findings indicated a significant increase in teacher candidates’ dispositions toward critical thinking after the intervention and a growth in their own learning. Moreover, Loes, Pascarella and Umbach (2012) investigated “… the unique effects of exposure to classroom diversity and involvement in interactional diversity on growth in critical thinking skills during the first year of college.” (p. 1). The researchers found that interactional diversity had a positive influence on critical thinking skills of diverse students.

Perkins and Murphy (2006) conducted an exploratory case study involving the development of a model for identifying and measuring individual engagement in critical thinking in an online asynchronous discussion, and underlined the potential usefulness and importance of identifying critical thinking in online asynchronous discussion groups based on their findings. Another researcher, Jeong (2003), examined group interaction and critical thinking in online threaded discussions. The researcher identified patterns in interactions and determined which interactions promoted critical thinking, and concluded that interactions having contradictory viewpoints stimulated more discussion and critical thinking. Furthermore, Walker (2004) examined the types of moves and strategies used by tutors facilitating the synchronous computer mediated communication debates in proportion to student responses in order to evaluate the efficacy of different move types. The researcher found that the most common move types were meta-statements, probe, challenge, inform and encourage

Based on the current literature, the findings of research studies are promising in terms of enhancing students’ critical thinking skills in online environments based on various approaches. Results also support a gain in students’ critical thinking dispositions.

**Instructional techniques applied in online environments**

Instructional techniques are educational activities which are shaped by instructional context like learning outcomes, content, and properties of a target group. Use of instructional techniques in online environments are so crucial since “… it is planned instructional methods that define formal education and allow for distinctions between serendipitous ‘web surfing’ and distance education” as stated by Kanuka, Rourke and Laflamme (2007, p. 261). Owing to this fact, many researchers investigated the impact of various instructional strategies like debate, Socratic questioning and problem-based learning, by considering different variables like academic achievement, critical learning, quality of instruction and deep learning (Kanuka, Rourke & Laflamme, 2007; Richardson & Ice, 2010; Khoshneshin, 2011; Park et. al., 2013; Lang et. al., 2013).

In their study, Şendağ and Odabaşı (2009) investigated how the online problem based learning (PBL) approach employed in an online learning environment influenced undergraduate students’ critical thinking skills (CTS) and content knowledge acquisition. The researchers concluded that learning in the online PBL group had a significant effect on increasing the critical thinking skills. Another researcher, Hou (2011), conducted a case study which empirically explored the learning process of adopting collaborative online instructional discussion activities for the purpose of problem-solving using situated scenarios in a higher education course. Based on the findings, the researcher suggested “… when compared to general situated learning activity, discussions are of better quality when they involve a role-playing activity, which also yields the most diverse options for solutions’” (p. 712).
Yang, Newby and Bill (2005) investigated the effects of using Socratic questioning to enhance students’ critical thinking (CT) skills in asynchronous discussion forums in university-level distance learning courses. The results of their study indicated that “… indicate that with appropriate course design and instructional interventions, CT skills can be cultivated and maintained in ADF [asynchronous discussion forums].” (p. 179). Similarly, Yang (2008), conducted a study to investigate whether students’ critical thinking skills would improve after they participated in Socratic dialogues asynchronous online discussion forums, and concluded that an inspired instructor and some energetic teaching assistants who use Socratic dialogues during small-group online discussions can successfully develop students’ critical thinking skills in a large university class.

Koh, Herring and Hew (2010) analysed the relationship between students’ levels of knowledge construction during asynchronous online discussions with respect to engagement in project-based learning. Researchers found that although instructor’s teaching discourse remained fairly consistent during project-based and non-project learning, students’ “… online discussions during project-based learning were characterised by more advanced levels of knowledge construction, where ideas were rationalised and integrated into plausible solutions.” (p. 284).

Kanuka, Rourke and Laflamme (2007) examined the influence of five groups of communication activities on the quality of students’ contributions to online discussion. The researchers considered nominal group technique, debate, invited expert, WebQuest and reflective deliberation as communication activities, and cognitive presence for specifying the quality of discussions. As a conclusion, the researchers suggested that instructional methods affect the quality of students’ contributions to online discussion. Richardson and Ice (2010) also investigated how various strategies can impact students’ critical thinking levels. The researchers considered a case-based discussion, a debate, and an open-ended discussion as instructional strategies and they reached an evidence of critical thinking and underline the importance of students’ comfort levels as a crucial factor in effective use of online discussions for enhancing higher order thinking skills.

**Background and importance of the study**

Although discussions are commonly used in online environments, pedagogical approaches used within these processes are rarely investigated. As stated by Bonk (2002), the major goal of more active and engagement for online learning experiences is to integrate expertise and experience of the learners to a group problem situation for discussion. Bonk (2002) also underlines the importance of usage of “… interactive and collaborative activities, a sense of variety and novelty in activities and delivery format, a sense of curiosity and fun in activities, engaging in discussion that involves multiple participants, and a supportive community of e-learners” (p. 90). Interesting and authentic problem situations from real context might increase learners’ critical thinking dispositions. Based on the fact that critical thinking dispositions are directly related with motivation, all the factors which can be used to increase motivation should be taken into consideration. Besides, by designing the instructional process with given special emphasis to some issues such as instructional techniques, discussion rules and assessment criteria, educators should provide the sustainability of the motivation. According to Walker (2005), strategic concerns like presence of subject matter experts and guest facilitators, use of variety of writing activities, convergent, divergent, evaluative and Socratic-questioning strategies, case studies, and role playing activities can encourage critical dialogue. Hence, it is obvious that integrating various instructional methods of learning and motivation to promote critical thinking in online discussion environments are very important for effective discussion processes.

All of these studies and suggestions reveal the impact of different implementations in different settings on the critical thinking skills of students, which is at the very least, evidence of enhancement of critical thinking skills in online environments by selecting various instructional techniques. Based on these facts, the aim of this research study was to explore the effect of instructional techniques on critical thinking skills and critical thinking dispositions of students in online discussion based environments. For achieving this goal, Six thinking hats, Brainstorming, Role playing, Socratic seminar, and Anyone here an expert, were considered as instructional techniques for implementation in online discussions.
Methodology

Research design

This study was carried out in triangulation design. “The Triangulation Design is a one-phase design in which researchers implement the quantitative and qualitative methods during the same timeframe and with equal weight. This design is used when a researcher wants to directly compare and contrast quantitative statistical results with qualitative findings or to validate or expand quantitative results with qualitative data” (Creswell and Clark, 2007, p. 62).

The quantitative part was conducted in pre-test/post-test comparison design of quasi-experimental design. The independent variable of the study was instructional techniques having five levels as Six thinking hats, Brainstorming, Role playing, Socratic seminar, and Anyone here an expert. The dependent variables of the study were critical thinking and critical thinking dispositions. The scores of the critical thinking dispositions were gathered through California Critical Thinking Disposition Inventory (CCTDI). In the qualitative part of the research, after the four-week experimental process, the messages in online discussion were analysed and digitised based on the content analysis model of critical thinking. The design is presented in Table 1.

Table 1. Design of the study

<table>
<thead>
<tr>
<th>Discussion groups</th>
<th>Pre-test</th>
<th>Discussion (4 weeks)</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Six thinking hats</td>
<td>CCTDI</td>
<td>D1</td>
<td>CCTDI Content analysis of critical thinking in online discussion (CACTOD)</td>
</tr>
<tr>
<td>Brainstorming</td>
<td>CCTDI</td>
<td>D2</td>
<td>CCTDI CACTOD</td>
</tr>
<tr>
<td>Role playing</td>
<td>CCTDI</td>
<td>D3</td>
<td>CCTDI CACTOD</td>
</tr>
<tr>
<td>Socratic seminar</td>
<td>CCTDI</td>
<td>D4</td>
<td>CCTDI CACTOD</td>
</tr>
<tr>
<td>Anyone here an expert</td>
<td>CCTDI</td>
<td>D5</td>
<td>CCTDI CACTOD</td>
</tr>
<tr>
<td>Mixed techniques</td>
<td>CCTDI</td>
<td>D6</td>
<td>CCTDI CACTOD</td>
</tr>
</tbody>
</table>

Participants

24 pre-service teachers who were attending a compulsory undergraduate course, the “distance education” offered by Computer Education and Instructional Technology (CEIT) at a private university in Turkey, participated in the study. Of the 24 students, 5 were feminine and 19 were male. Participants were randomly assigned to one of six groups by trying to equate academic achievement levels of group members (i.e. we tried to form homogenous groups). Each group composed of 4 students and there were a total of six groups formed for this study.

Context and process

The “Foundations of Distance Education” course was designed in a blended way, which uses both traditional and online activities through a learning management system, in concert. Chat was used as an online discussion tool by students to direct discussion sessions. During discussion processes, some students used the chat tool provided within the learning management system, while others preferred to chat via MSN messenger since they felt more comfortable with this system due to their prior experience. Starting with the 10th week of the course, chat sessions were conducted with six groups having different moderators each week. Hence, each student in each group played the role of moderator once. During discussions, Six thinking hats, Brainstorming, Role playing, Socratic seminar, and
Anyone here an expert, were used as instructional techniques for constructing the discussion process. The tutor didn’t participate in any of the discussion sessions in order to prevent bias, but all the necessary information needed for an effective discussion process like ethics, discussion rules, implementation ideas and details of instructional technique were provided before discussion sessions.

With the Six thinking hats instructional techniques, students share their ideas about the problem from different perspectives according to the hat they are wearing. Parallel with the name of this technique, there are six hats with colours of white, red, black, yellow, green and blue. When students prefer the white hat, they might ask questions or call for information about the problem. With the red hat, they might express their emotions; with black hat, they might judge; with yellow hat, might be optimistic; with green hat, they might be creative including possibilities, alternatives or new ideas and with blue hat, they might think about their thinking process. While adopting this technique into online discussion, students wrote their thoughts with a different font colour according to the hats’ meaning.

Brainstorming encourages students to share their ideas freely about the topic or problem. In this technique, students can express themselves with relevant or irrelevant words or ideas that are accepted without any criticism. After the idea sharing process is completed, teachers can go through the results and evaluate the responses for answering the problem. While adopting this technique into online discussion, students wrote relevant or irrelevant words or ideas. Then moderators evaluated the responses for answering the problem.

Role playing encourages students generate their ideas about the problem from different perspectives according to the roles they are assigned. With this technique, students might solve problem or share ideas from another perspective. In this study, students discussed their problems with assigned roles such as administrator, teacher, student and content specialist. They wrote assigned roles in parentheses with their nicknames.

Socratic seminar guides students to generate their ideas by asking questions of them, requesting clarification, evidence or suggestions from them. These interactive dialogues foster students to think more critically.

Anyone here an expert is similar to role playing. This technique encourages students to generate their ideas about the problem from different specialties they are assigned. In this study, students discussed their problems with assigned specialties such as technologist, programmer, social network expert and instructional designer.

Mixed techniques are composed of techniques mentioned above. Each week students discussed with different techniques respectively.

The students in this study have enrolled on a compulsory course “special instructional methods” focusing on the theoretical and practical issues of instructional methods and techniques. Nevertheless, instructional techniques were explained both in writing and verbally to the students. Before discussion sessions, California Critical Thinking Disposition Inventory was implemented for the students. Then the discussion topics were announced to students in order to provide students with enough time to search and get prepared for the session. After four weeks CCTDI were implemented to the students as a post-test.

The topics selected for discussion were as follows:
Do some professions such as Doctor, nurse, or paramedic, have their education in distance education programs? How much you can rely on these people when they receive a diploma from distance education programs?

We are aware that virtual communities create new structures based on a social software (blog, wiki, forum, chat, etc.) and social network (Facebook, Twitter, Flickr, Grou.ps, Ning, Delicious, etc.) for knowledge sharing. In this framework, discuss what are these applications? And which purposes are they used for?
Which teaching instructional methods and techniques may be more appropriate for e-learning process?
What do you foresee in terms of changes and innovations for the future of distance education? Based on your discussion what are your perceptions about the possible scenarios that we may face in the next five years?
Research questions

The aim of this research study was to explore the effect of instructional techniques on critical thinking and critical thinking dispositions in online discussion. For this purpose, questions presented below were answered:
1. Do instructional techniques affect critical thinking dispositions?
2. Do instructional techniques affect critical thinking?
3. What were pre-service teachers’ critical thinking and critical thinking dispositions?

Data collection techniques

Quantitative - California critical thinking disposition inventory (CCTDI)

The original CCTDI includes 75 items having seven sub-scales. These are truth-seeking, open-mindedness, analyticity, systematicity, self-confidence, inquisitiveness and maturity. Kökdemir (2003) adapted this inventory to the Turkish. According to this adaptation study, 51 items having six sub-scales, analyticity, open-mindedness, inquisitiveness, self-confidence, truth-seeking, systematicity, were kept in the scale. The reliability coefficients of each sub-scale ranged from .61 to .78. Reliability of the whole scale was found to be .88 (Kökdemir, 2003).

Qualitative - Content analysis of the online discussion

Garrison, Anderson, and Archer (2001) created an efficient and reliable electronic assessment tool for the critical-thinking process (i.e., cognitive presence) as reflected in a computer-conference transcript. The element of cognitive presence has four categories in coding template such as triggering event, exploration, integration connecting ideas and resolution.

The first phase of the model is considered as the triggering event. In this event, an issue, dilemma, or problem that emerges from experience is identified or recognised. In the second phase, exploration, participants shift between the private, reflective world of the individual and the social exploration of ideas. This is a phase that students brainstorm, question, and exchange of information. In the third phase, integration, students construct meaning from the ideas generated in the previous phase. The last phase is a resolution of the dilemma or problem by means of direct or vicarious action.

Data analysis

To reveal whether there is a difference between groups, two way variance analysis (ANOVA) was calculated with the data gathered from Quantitative - California Critical Thinking Disposition Inventory. Moreover, the data in the online discussion was analysed and digitised based on the content analysis model of Garrison, Anderson, and Archer (2001).

Results

Critical thinking dispositions of pre-service teachers

The descriptive statistics of critical thinking dispositions of pre-service teachers are presented in Table 2.

<table>
<thead>
<tr>
<th>Instructional Techniques</th>
<th>Pre-Test</th>
<th>Post-Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
</tr>
<tr>
<td>Six thinking hats</td>
<td>4</td>
<td>205,75</td>
</tr>
<tr>
<td>Brainstorming</td>
<td>4</td>
<td>207,50</td>
</tr>
<tr>
<td>Role playing</td>
<td>4</td>
<td>201,00</td>
</tr>
<tr>
<td>Socratic seminar</td>
<td>4</td>
<td>195,50</td>
</tr>
</tbody>
</table>
As seen in Table 2, the mean difference of pre- and post-test results of the groups showed very little difference. The major difference was observed for Socratic Seminar group. Therefore, the ANOVA results tested for whether there was a significant difference between pre- and post-test results (see Table 3). According to the results, there wasn’t a significant difference between pre- and post-test results of the students’ critical thinking dispositions [F(1,18)=0.47 p>.05]. In other words, discussing with different instructional techniques did not cause any significant differences to critical thinking dispositions. This result may show that the instructional techniques may not have impact on critical thinking dispositions.

**Table 3. The ANOVA results**

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between-Subjects</td>
<td>12336,667</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>913,167</td>
<td>5</td>
<td>182,633</td>
<td>.288</td>
<td>.914</td>
</tr>
<tr>
<td>Error</td>
<td>11423,500</td>
<td>18</td>
<td>634,639</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within-Subjects</td>
<td>6313,000</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-post tests</td>
<td>14,083</td>
<td>1</td>
<td>14,083</td>
<td>.046</td>
<td>.833</td>
</tr>
<tr>
<td>Group*Pre-post tests</td>
<td>732,917</td>
<td>5</td>
<td>146,583</td>
<td>.474</td>
<td>.791</td>
</tr>
<tr>
<td>Error</td>
<td>5566,000</td>
<td>18</td>
<td>309,222</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>18649,667</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

When weekly participation of pre-service teachers to online discussion according to instructional techniques was considered (Table 4), for the Six thinking hats group, the total number of sentences was 878, while the number of coded sentences was 473 and the number of non-coded sentences was 405. For the Brainstorming group, the total number of sentences was 558, while the number of coded sentences was 474 and the number of non-coded sentences was 84. For the Role playing group, the total number of sentences was 391, while the number of coded sentences was 272 and the number of non-coded sentences was 119. For the Socratic seminar group, the total number of sentences was 453, while the number of coded sentences was 393 and the number of non-coded sentences was 60. For the Anyone here an expert group, the total number of sentences was 663, while the number of coded sentences was 517 and the number of non-coded sentences was 146. For the Mixed techniques group, the total number of sentences was 603, while the number of coded sentences was 519 and the number of non-coded sentences was 84.

**Table 4. The number of coded, non-coded and total sentences**

<table>
<thead>
<tr>
<th>Instructional Technique</th>
<th>Number of Coded Sentences</th>
<th>Number of Non-Coded Sentences</th>
<th>Number of Total Sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Six thinking hats</td>
<td>473</td>
<td>405</td>
<td>878</td>
</tr>
<tr>
<td>Brainstorming</td>
<td>474</td>
<td>84</td>
<td>558</td>
</tr>
<tr>
<td>Role playing</td>
<td>272</td>
<td>119</td>
<td>391</td>
</tr>
<tr>
<td>Socratic seminar</td>
<td>393</td>
<td>60</td>
<td>453</td>
</tr>
<tr>
<td>Anyone here an expert</td>
<td>517</td>
<td>146</td>
<td>663</td>
</tr>
<tr>
<td>Mixed techniques</td>
<td>519</td>
<td>84</td>
<td>603</td>
</tr>
</tbody>
</table>

When the data of weekly participation of pre-service teachers to online discussion according to instructional techniques was examined in general, it is easily seen that pre-service students shared the highest number of sentences in the Six thinking hats group. In second place was the Anyone here an expert group, and the third was Mixed techniques. But when coded sentences were examined, the Mixed techniques group was first, the Anyone here an expert group was second and the Brainstorming group was third. These results may show that the Mixed techniques group performed with the best ability of critical thinking, the Anyone here an expert group was second and the Brainstorming group was third in terms of performing critical thinking ability in online discussion.
Critical thinking of pre-service teachers in online discussion

When the phases of the critical thinking were examined in the coded sentences (Figure 1), the Six thinking hats group has 78 sentences for triggering event, 315 for exploration, 92 for integration and 0 for resolution phase. The Brainstorming group has 37 sentences for triggering event, 333 for exploration, 107 for integration and 0 for resolution phase. The Role playing group has 52 sentences for triggering event, 193 for exploration, 30 for integration and 0 for resolution phase. The Socratic seminar group has 61 sentences for triggering event, 297 for exploration, 42 for integration and 0 for resolution phase. The Anyone here an expert group has 81 sentences for triggering event, 409 for exploration, 39 for integration and 0 for resolution phase. And finally, the Mixed techniques group has 61 sentences for triggering event, 397 for exploration, 90 for integration and 0 for resolution phase.

In the triggering event phase, the group of Anyone here an expert shared the most ideas, the second was Six thinking hats and joint third was Mixed techniques and Socratic seminar. In the exploration phase, the group of Anyone here an expert shared the most ideas, the second was Mixed techniques and Brainstorming was third. In the integration phase, the group of Brainstorming shared the most ideas, the second was Six thinking hats and the third was Mixed techniques. Finally, no group shared ideas matching with resolution phase. Sample statements from different discussion groups are presented below (Table 5).

<table>
<thead>
<tr>
<th>Phases</th>
<th>Sample Statements from different discussion groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triggering event</td>
<td>Do you have any suggestions as a methodology expert? Which software should be used? (Anyone here an expert)</td>
</tr>
<tr>
<td></td>
<td>To what extend can you trust doctors in case they educated at a distance and got their diplomas based on a distance education model? (Socrates seminar)</td>
</tr>
<tr>
<td>Exploration</td>
<td>I think that both educators and students have the chance of learning via interactive technologies like whiteboard and audio-conference applications (Mixed techniques)</td>
</tr>
<tr>
<td></td>
<td>According to me that should be robots donated with human characteristics including the body infrastructure. These robots can be used by doctoral candidates to make their applications. (Brainstorming)</td>
</tr>
<tr>
<td></td>
<td>Students may prefer more instructor-controlled environments for their teaching-learning processes. From different points of view, each social network can provide various contributions, but age and student characteristics are so important (Anyone here an expert)</td>
</tr>
<tr>
<td></td>
<td>I prefer demonstration and hands-on experience in general. The method that I will choose can also change according to the course content. Let me introduce you the demonstration method (Role Playing)</td>
</tr>
<tr>
<td>Integration</td>
<td>I also agree that we cannot trust to all information shared on the Internet (Anyone here an expert)</td>
</tr>
</tbody>
</table>
expert)

Yes, we can get members of various social network sites and make applications by using only one web address (Mixed techniques)

Let us summarise what we have talked about up to now. Both face-to-face education and distance education has advantages and disadvantages. The important point is to merge the possible advantages of the two methods to reveal an effective approach by eliminating disadvantages (Six Thinking Hats)

This is really a very nice innovation. As our moderator underlined, we currently cannot prevent cheating in distance education exams. By using this new technology, a new era is beginning. (Role Playing)

Resolution

None coded

Discussion and conclusion

The aim of this research study was to explore the effect of instructional techniques on critical thinking and critical thinking dispositions in online discussion. For exploring this effect, Six thinking hats, Brainstorming, Role playing, Socratic seminar, and Anyone here an expert were selected as an instructional techniques for online discussion.

Based on the quantitative analysis, it can be concluded from this research study that discussing with different instructional techniques (Six thinking hats, Brainstorming, Role playing, Socratic seminar, Anyone here an expert) did not cause any significant differences to critical thinking dispositions. The reason for this finding may be related with the motivation levels of the students, since critical thinking disposition of the students is directly related with their personal characteristics. On the other hand, the low number of students (each group was composed of 4 students) and the shortness of the discussion period (4 weeks) might also be indicated as the possible reasons for this finding.

Some other possible contributors to this finding may be that students were not familiar with the instructional techniques (in fact this was the first time for an implementation of an instructional technique and discussing for a course in a virtual environment for all students) and this virtual discussion application was the first experience of students in terms of being a moderator. Some suggestions for overcoming these possible obstacles may be conducting discussions in larger groups (about 8-10 students) and making an increase in the duration of the discussion intervals (at least 6 or 8 weeks). Parallel with these ideas, exploring the process in depth by observing and making interviews may reveal obstacles faced by the students and give punctual and more realistic information about the context.

Another finding is that although not statistically significant, there is at least a difference between the Socratic seminar group and the other groups. Maybe in the discussion process, Socratic questioning helped students to exchange their thoughts, evaluate their perspectives critically and come to the conclusion about the discussion topic. This finding shows similarity with the other findings reported in the literature. For example, results of the study of Yang, Newby and Bill (2005) indicated that Socratic questioning helped students demonstrate a higher level of CT skills. Another study of Yang (2006) resulted that critical thinking dispositions could be enhanced via the teaching and modelling of Socratic dialogues on a series of asynchronous online discussions.

When the qualitative part of the study was examined, the Socratic seminar group performed with the least ability of critical thinking. In the qualitative part, the Mixed techniques group performed with the best ability of critical thinking, the Anyone here an expert group was the second and the Brainstorming group was third in terms of performing critical thinking ability in online discussion. Getting the best performance from Mixed techniques groups can be interpreted as this group used all of the instructional techniques one by one. This experience probably made the group realise the strengths and weaknesses of the discussion process and thereby make differentiation between the techniques. This group should also be more motivated at each discussion process, since they will try another instructional technique which increases their anxiety more than other groups. Individual differences between students might also cause for this result. Every student can prefer another technique for discussing. Thus, if one student cannot show a good performance with one technique, he/she can contribute more with another. As also stated by Keller and Suzuki (2004), “No matter how interesting a given tactic is, people will adapt to it and lose interest over time. Thus, it is important to vary one’s approaches and introduce changes of pace” (p. 231). Hence, for discussion sessions in e-Learning environments the use of mixed techniques should be suggested for effectiveness in teaching. On the other hand, another way of increasing critical thinking performance and dispositions might be the presence of
the instructor in the discussion environment. Similarly, some studies conclude the importance of expert contribution to the discussion processes in terms of critical thinking (Yang, 2008; Dennen, 2002; Havard, Du & Olinzock, 2005). Hence a future study may apply these or other discussion techniques in online environments with the help of an expert.

The findings of the study indicated that Anyone here an expert group was the second in terms of critical thinking performance. Like the first group, this group also had some sources of motivation for each discussion session, since their roles of expertise and their viewpoints for discussing the topic changed each week. Considering the discussion topic from diverse expertise approaches, trying to get different point of views and evaluating new ideas from these diverse roles should lead each group member to express themselves in a more efficient way. Finally, getting Brainstorming group as the third group in critical thinking performance is not surprising, since this technique fosters creativity without worrying about the correctness of the proposed ideas in a flexible climate and is an unbounded discussion environment. Hence dealing with the motivation variable for the discussion processes can be another aspect for future studies.

Although limited in terms of the number of students, the duration of the discussions and the discussion topics, this research study revealed important aspects of virtual discussions in terms of critical thinking phenomenon. The main point which emerged from this study is the importance of using different instructional techniques or using instructional techniques that should make students think diversely for each distinct discussion process. Based on the findings of this study, it is obvious that when students perceive the discussion process as an ordinary situation, after each week they got used to it and their performance did not increase. As a conclusion, if the instructor makes at least slight differences in terms of nature of behaviour and thinking processes that the student will show in the discussion process, the performance of student’s increases in a very short time with very few people. As instructors we should provide our students with rich learning environments and a variety of learning possibilities for effective teaching which shows the importance of considering individual differences and diversity in our instructional design processes (Bonk, 2002; Walker, 2005). Thus, future research studies should reconsider similar discussion topics or dilemma, and make use of different instructional techniques in order to measure critical thinking and dispositions by changing the duration and number of participants.

References


Adaptive Instruction to Learner Expertise with Bimodal Process-oriented Worked-out Examples

Jihyun Si, Dongsik Kim* and Chungsoo Na
Hanyang University, Dept. of Educational Technology, 17 Haengdang-dong, Seongdong-gu, Seoul Korea, 133-791 // jennyhan0306@gmail.com // kimsik@hanyang.ac.kr // skcjdtn@gmail.com

*Corresponding author

(Submitted December 6, 2012; Revised March 8, 2013; Accepted May 7, 2013)

ABSTRACT
This study investigated the instructional efficiency of adaptive instruction to learner expertise in the domain of C programming language with college students. It also aimed to investigate whether a bimodal process-oriented worked-out example (WOE) could effectively control extraneous cognitive load and further improve instructional efficiency. For this purpose, a learner-paced problem-solving e-learning environment was developed. A total of 112 college students participated and they were randomly divided into four groups (adaptive and bimodal, adaptive and unimodal, fixed and bimodal, and fixed and unimodal) when they logged into the problem-solving e-learning environment. After removing uncompleted or repeated data, data from 96 students were used for a series of ANOVA and ANCOVA. The findings showed that the adaptive instruction groups showed significantly higher instructional efficiency than the fixed instruction groups. Although there was no significant difference between the bimodal and the unimodal WOE groups, the bimodal WOE groups showed lower mental effort, higher knowledge acquisition, and instructional efficiency. In addition, the bimodal WOE condition was more efficient in both adaptive and fixed conditions. Based on these findings, it was concluded that the adaptive instruction method and the bimodal process-oriented WOE effectively controlled cognitive load and thereby successfully and efficiently led to schema construction and automation. The insight gained through this study may inform instructional designers seeking to enhance their understanding of efficient instructional design within the cognitive load theory framework.

Keywords
Adaptive instruction, Cognitive load theory, Process-oriented WOE, Instructional efficiency, E-learning, Learner expertise, Modality effect

Introduction
Adaptive instruction refers to educational interventions intended to effectively accommodate the needs of individual students (Park & Lee, 2003). This educational approach is generally characterized as one that incorporates alternative instructional procedures by providing built-in flexibility to allow students to take various paths to learning (Park & Lee, 2003). The practice of adaptive instructions has had a long history (e.g., aptitude-treatment interactions, Cronbach & Snow, 1977). More recently, researchers in the cognitive load theory (CLT) have investigated methods of adaptive instructions and cast new insight on this important issue.

CLT offers principles and methods to design and deliver instructional interventions that efficiently utilize the limited capacity of human working memory (Sweller, van Merriënboer, & Paas, 1998). If instructional interventions are designed in a way that causes excessive cognitive load, the limited capacity of human working memory can easily be overloaded. This excessively high cognitive load does not contribute to acquisition or automation of schematic knowledge, but rather impedes it (Paas, Renkl, & Sweller, 2004). One of the instructional methods to control excessive cognitive load is worked-out examples (WOE) (Renkl & Atkinson, 2010). WOE provides an experts’ problem solving model for novices to study and emulate, so a substantial amount of unnecessary cognitive load caused by premature problem solving can be reduced by studying with WOE in the early phase of skill acquisition (Sweller, 2006). However, as learner expertise grows, knowledge acquisition from studying WOE becomes a redundant activity that contributes little or nothing to further learning. Instead, problem solving fosters learning more than studying with WOE (Kalyuga, Chandler, Tuovinen, & Sweller, 2001). This reversal of the WOE effect is called the expertise reversal effect (Kalyuga, Ayres, Chandler, & Sweller, 2003). The instructional implication of this expertise reversal effect is continuous optimization of cognitive load.
Adaptive instruction to learner expertise

To continuously optimize cognitive load, instructional designs need be tailored to an individual trajectory of cognitive skill acquisition in a domain (Kalyuga, 2007, 2008; Kalyuga & Sweller, 2004, 2005). In the traditional learning situations, researchers or instructors pick when to modify instructional techniques typically through interviews or think-aloud procedures or observations as learners gain expertise. However, in e-learning environments, such techniques are inconvenient to use (Kalyuga, 2006a). As an alternative, CLT researchers have utilized efficiency measures to decide the right movement. Such measures were originally developed to measure efficiency of instructional conditions. Pass and van Merriënboer (1993) suggested the following formula to calculate instructional efficiency, $E = \frac{Z_{\text{performance}} - Z_{\text{mental effort}}}{\sqrt{2}}$. Students’ performance and mental effort on the test are first standardized and then mean standardized test performance and test mental effort are entered into this formula to calculate instructional efficiency. According to them, a high performance combined with low mental efforts is called high instructional efficiency while a low performance combined with high mental effort is called low instructional efficiency. However, a combination of performance and mental effort is also indicative of expertise. Learners with more expertise are able to attain equal or higher levels of performance with less investment of mental effort (Kalyuga, 2007). Thus, the efficiency measure is also used to measure levels of learner expertise (van Gog & Paas, 2008).

However, such formula may not be appropriate for e-learning situations. Kalyuga and Sweller (2005) argued that it is not convenient for e-learning applications which require diagnostic assessments of learner expertise for adaptation in real time during an experiment. They defined efficiency as the ratio of the level of performance to rating of mental effort, $E = P/R$ ($P$: performance measure, $R$: mental effort rating). They also employed a critical level of efficiency defined as $E_{\text{cr}} = P_{\text{max}} / 9$ (as 9-point subjective rating scale was used to assess cognitive load), where $P_{\text{max}}$ is the maximum performance score for a given task level. If the efficiency measure is $E > E_{\text{cr}}$, the cognitive performance is considered efficient. If $E \leq E_{\text{cr}}$, the cognitive performance is regarded as relatively inefficient. Similarly, Kalyuga (2006a) used a simple threshold-based definition for efficiency. In this definition, learners’ performance was considered efficient if they correctly verified all the suggested solution steps up to, but not including, the final numerical answer, and rated a task difficulty as less than 5 on 9-point rating scales. Although these formulas to calculate efficiency are slightly different, they indicates that if similar levels of performance are reached with less mental effort, the efficiency is higher and they are convenient to use in e-learning environments.

Based on the definition of efficiency described above, Kalyuga (2006a) and Kalyuga and Sweller (2004, 2005) investigated the efficiency of adaptive instructions to learner expertise in e-learning environments. They all used a subjective rating of mental effort to measure cognitive load (Paas, & van Merriënboer, 1993). To measure learner performance, two different diagnostic methods were employed, the first-step method (Kalyuga, 2006c; Kalyuga & Sweller, 2004, 2005) and rapid verification method (Kalyuga, 2006a, 2006b). The rationale behind both methods was drawn from a theory of long-term working memory. According to the theory of long-term working memory, “reliance on acquired memory skills enables individuals to use long-term memory as an efficient extension of working memory in particular domains and activities after sufficient practice and training” (Ericsson & Kintsch, 1995, p. 211). That is, when learners encounter a problem in a familiar domain, their available knowledge structures in long-term memory are rapidly activated and corresponding long-term working memory structures created. These schematic knowledge structures are durable and interference-proof enough to observe their immediate traces (Kalyuga, & Sweller, 2004). Such schematic knowledge structures provide necessary executive guidance during problem solving and they are a major feature of learner expertise. (Kalyuga, 2006c, Sweller, 2006, Ericsson & Kintsch, 1995). To probe this knowledge structure, the first-step method asks learners to indicate the first-step toward a solution of a task for a limited time while the rapid verification method asks them to verify rapidly each of the sequentially presented solution steps. If a task is precisely specified in advance, the first step methods may be utilized. However, when the solution procedure requires drawing graphical representations or there are several possible solution paths but only a limited number of steps representing different levels of schema-based problem solutions can be selected, the level of proficiency can be assessed by the rapid verification method. The studies of Kalyuga (2006a, 2006b, 2006c) and Kalyuga and Sweller (2004, 2005) showed that both methods have a sufficiently high degree of external validity to detect differences instantly and adequately in learner knowledge structures.

In both the studies of Kalyuga (2006a) and Kalyuga and Sweller (2005), adaptation occurred based on the efficiency measures described before. The decisions for the initial placements for the training and the adjustment of levels of
instructional guidance during the training were made according to the efficiency measures. Kalyuga and Sweller study (2005) compared the learner-adapted instruction with the non-adapted instruction in the domain of 10th grade algebra. According to their results, the learner-adapted group showed significantly higher efficiency gains and knowledge gains than the control group. With these results, they concluded that the learner-adapted instruction through the first-step method coupled with mental effort ratings proved to be more efficient than non-adapted instruction. Similarly, Kalyuga (2006a) compared two learner-adapted instructions (efficiency-based and performance-based) to non-adapted instruction with vector addition motion problems in the domain of 11th grade kinematics. The results showed that both of these adaptive groups through a verification method coupled with mental effort ratings outperformed the non-adapted group on knowledge gains, spent less mental effort and less instruction time, and showed higher instructional efficiency.

**Bimodal process-oriented WOE**

To reduce unnecessary cognitive load imposed on novices, Kalyuga (2006a) and Kalyuga and Sweller (2005) used a typical WOE as an instructional intervention. A typical or product-oriented WOE provides a problem state, solution steps, and a goal state but does not include the rationale for the solution steps (van Gog, Paas, & van Merriënboer, 2004). In contrast, a process-oriented WOE states not only the solution steps, but also the rationale behind those steps. That is, it includes the “why” information in order to explain functions of operators and the “how” information in order to inform learners of the strategic knowledge used by experts. Such information is useful particularly at the initial training stage when students lack domain knowledge. It increases learners’ understanding of problem solving and encourages schema construction and automation (van Gog et al., 2004).

In designing a process-oriented WOE, the modality effect needs to be considered; otherwise, the effectiveness of learning from the process-oriented WOE could be moderated by inefficient design of WOE. According to Baddeley and Hitch (1974), working memory has two components, the phonological loop and the visual-spatial sketchpad and these two channels partly independently process auditory or visual information. Thus, if the process-oriented WOE is presented as a bimodal WOE, by allocating “why” and “how” information into auditory and visual channels rather than dedicating them to one channel, the learner is able to off-load some of the cognitive processing from the overloaded visual channel to the not-overloaded verbal channel (Mayer & Moreno, 2010). Many empirical studies have provided supports for the superiority of multimodal presentation of information (Moreno & Mayer, 2002; Mousavi, Low, & Sweller, 1995). However, Tabbers, Martens, and van Merriënboer (2004) have shown different results that replacing visual text with spoken text does not always improve learning; they found no modality effect. Tabbers et al. (2004) argued that in a learner-paced system, the visual-only condition was more effective as the learners with the visual-only instructions had more time to relate the text to the picture and could jump back and forth through the text more easily than with spoken text instruction. These mixed results need to be clarified.

**The purpose of the study**

Kalyuga (2006a), and Kalyuga and Sweller (2005) showed the possibility and promising potential of adaptive instructions to learner expertise in e-learning environments, but there is lack of research available to confirm and generalize their findings. Further research in other areas, in particular, complex cognitive skill acquisition, is necessary to generalize their findings. Typical complex cognitive skills have both recurrent and non-recurrent components (van Gog et al., 2004). The solution steps of complex problems are not easy to be specified in advance as recurrent algorithmic skills have a narrow problem space but non-recurrent skills have multiple possible solution paths (van Gog et al., 2004). In these cases, a rapid verification method may be more appropriate to probe changing levels of learner expertise in adaptive e-learning environments.

Their adaptive instructions could be further improved with a bimodal process-oriented WOE. If problems under instructions are complex, a typical WOE is not enough to facilitate schema construction and automation. In contrast, a bimodal process-oriented WOE provides necessary information for schema construction and automation in a cognitively efficient way. According to the modality effect, the utilization of the bimodal process-orientated WOE is expected to effectively expand working memory capacity, thereby resulting in learning enhancement. However, the utilization of the bimodal process-oriented WOE has not been investigated in adaptive instructions in e-learning environments. Furthermore, there is a study reporting that the modality effect was not found in a learner-paced
system as opposed to other studies supporting the modality effect. Theses mixed results need be clarified. Therefore, the purpose of this experimental study is to test the efficiency of adaptive instruction to learner expertise based on efficiency measures in a complex problem-solving e-learning environment. The rapid verification method was utilized for cognitive diagnosis coupled with a subjective mental effort rating to calculate the efficiency measures. In addition, we investigated whether a bimodal process-oriented WOE could effectively control cognitive load and further improve instructional efficiency in the learner-paced e-learning environments. The specific research question for this purpose was “How do the different instruction methods (adaptive and fixed instruction) and the modality of WOE (bimodal and unimodal) affect participants’ knowledge acquisition, mental effort, and instructional efficiency?”

Method

Participants

One hundred and twelve students who enrolled in one of the four sessions of basic computer programing classes at a four-year teachers college in 2012 spring semester were invited to participate in this study. One instructor taught all the classes. When 112 students logged on to the website developed for this experiment, the system randomly assigned them into one of the four groups, adaptive instruction and bimodal group (AB), adaptive instruction and unimodal group (AU), fixed instruction and bimodal group (FB), and fixed instruction and unimodal group (FU). Each experimental group was expected to have an equal sample size, but random loss of a few subjects caused unequal sample size; 5 students did not complete the experiment, and 9 students participated more than once. Thus, 2 students, one from the AB group and one from the FU group, were randomly deleted to make all the four groups have equal samples. Thus, data from 96 students were prepared for statistical analyses (24 for each group, 71 female, 25 male). Their age ranged from 20 to 39 (M = 23.68) and they were all juniors. Their majors ranged from Music, Mathematics, Science, Physical Education, to Practical Art.

Experimental materials

Task classes and sub-steps of problems

The content area in this experiment was C programming language. The desired exit-behavior of learning was for students to be able to write a basic source program with C programming language appropriately. This complex cognitive skill has both recurrent and non-recurrent components. Since the participants were novices in this area, the overall complexity of the learning materials was adjusted to a very basic level. There were three complexity levels (task classes), if-else statement, if-else compound statement, and while statement with if-else statement. The problems in each task class had three sub-steps: verify an appropriate flow chart; declare variables; and express if-else or if-else compound, or while statements based on the flow chart chosen.

Bimodal and unimodal process-oriented WOE

Two types of process-oriented WOE were constructed, bimodal process-oriented WOE (visual and auditory mode) and unimodal process-oriented WOE (visual-only mode). For the unimodal WOE, all the information including the rationale of the solution steps was visually presented (see Figure 1). For the bimodal WOE, as seen in Figure 2, the rationale coordinated with the flow chart of the unimodal WOE was removed and instead, a control box was presented to play an audio file. Both WOE have three sub-steps and each sub-step was used as a heading as seen in Figure 1and 2. Both the WOE presented all the sub-steps at the beginning of the training session and for smooth transitions from learning from the WOE to problem solving, rather than abrupt changes in cognitive demand from WOE to the to-be-solved problems, their sub-steps were progressively faded backward (see Table 1). Two types of the process-oriented WOE occurred only in the training session. In the learning phase, all the participants studied only with the unimodal process-oriented WOE.
Table 1. Faded WOE during the training session

<table>
<thead>
<tr>
<th>Task Class1</th>
<th>Task Class2</th>
<th>Task Class3</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>1.1</td>
<td>2.1</td>
<td>3.1</td>
</tr>
<tr>
<td>W</td>
<td>W</td>
<td>W</td>
</tr>
<tr>
<td>Sub-step1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sub-step2</td>
<td>W</td>
<td>W</td>
</tr>
<tr>
<td>Sub-step3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. P: Problem  W: Worked-Out Example

Subjective mental effort

A 9-point Likert scale rating was utilized to measure mental effort. This subjective rating scale are based on the assumption that participants are able to reflect on and report their mental effort expenditure (Paas, Tuovinen, Tabbers, & van Gerven, 2003). The scale used in this study was the same form employed in the studies of Kalyuga (2006a), and Kalyuga and Sweller (2005). Learners were asked to rate their perceived mental effort with semantically different scales varying from extremely easy (1) to extremely difficult (9) (see Figure 3).

Problem solving e-learning environment

The problem solving e-learning environment was developed and its structure is presented in Figure 4. As seen in Figure 4, it contained pretests, learning phase, initial diagnostic tests, training sessions, mental effort rating for the
whole training session, and final diagnostic tests. The participants could access this learner-paced e-learning environment from IBM compatible PCs, anywhere if they have an Internet access. This whole learning package was developed by the author consulting leading C programming language books and reviewed thoroughly by several graduate students in the field of Computer Education. Then, the whole package was inspected by an instructor who has taught C programming language for the last three years. With this reviewed materials, this problem solving e-learning environment was developed using PowerPoint, HTML, PHP, MySQL, and JSP. Then, it was field-tested with a group of graduate students (N = 12). They were also asked to check clarity and wording of the whole package. Their feedback was incorporated and this e-learning environment was finalized.

![Training session diagram](image)

*Figure 4. Structure of the problem-solving e-learning environment*

*Login page*

The login page was developed to collect the participants’ background information such as name, student number, gender, age, major and experience with C programming language.

*Pretests*

The pretest was designed to assess participants’ prior knowledge in C programming language and consisted of 10 multiple-choice questions.

*Learning phase*

In this phase, the learners studied brief explanations of C programming language and if-else statement, if-else compound statement, and while statement with if-else. Each explanation of the three statements was followed with one WOE paired with one conventional problem to practice. All WOEs and conventional problems consisted of a sequence of three sub-steps. The participants were allowed to move back to the previous pages by clicking on the “Prev.” button in this learning phase.

*Initial diagnostic tests*

The initial diagnostic test consisted of three problems (one problem from each task class). Each was designed as a sequence of three sub-steps. The rapid verification method was employed to assess learner expertise as the task involved graphical information and included both non-recurrent and recurrent skills. A 9-point mental effort rating scale was presented to measure mental effort after the participants clicked on an answer for every question (see Figure 3). The first sub-step question for each problem provided a flow chart and asked the learners to choose one of the three options, “correct”, “incorrect” and “don’t knows” (see Figure 5). For the next two sub-steps, four multiple-choice recognition questions were utilized due to the technical difficulty to make verification questions with long source programs. They were asked to choose one of the four alternatives for the blanks in the problem (see Figure 6 and 7). These sub-steps were scored differently. For the first sub-step, score 1 was allocated for a correct response. For the second sub-step, score 2 was allocated and for the third sub-step, score 3 was allocated for a correct answer. As the second and the third sub-steps required the participants to maintain the results of the previous step in working memory to mentally determine the values of the current stage, their scores became cumulative. Thus, if a learner chose correct answers for all the sub-steps, the allocated maximum score was 6 (1+2+3 = 6).
Training session

The training session included nine stages and each of three stages corresponded to each task class. Each stage contained a fully or faded worked-out example, one conventional problem solving, one diagnostic test, and a mental effort rating (see Figure 8). In the fixed instruction groups, the participants followed the fixed sequence from stage 1 to stage 9 without any repetition. In contrast, in adaptive instruction groups, the initial placement for an appropriate
training stage occurred according to the efficiency measure based on the simple threshold-based definition discussed before (Kaluga, 2006a). The participants started the training session from the stage corresponding to either the first initial diagnostic question not answered correctly or the first mental effort rating of 5 or above in the initial diagnostic tests. For example, if a learner chose a wrong answer or rated mental effort as 5 or both for problem 1.2 (the second sub-step of problem1) in the initial diagnostic tests, she or he started the training from the stage 2 of the training session. During the training session, repetitions occurred based on the results of the diagnostic test taken at the end of each stage. If participants correctly answered a diagnostic test and rated the mental effort as 4 or below, they advanced to the next stage. However, if they did not select a correct answer, or rated mental effort as a 5 or above, or both, they studied the same stage again. After the second try, the participants advanced to the next stage regardless of their answer and mental effort rating. After checking an answer or rating mental effort, the learners received messages such as “Good job, click on the next button”, “Good job, but it seemed difficult for you. Study the next problem.” If they incorrectly answered the test on the second try regardless of their mental effort rating, a correct answer was provided with the following feedback, “Incorrect. The right answer is _____ . Study the next example.” The fixed instruction groups also took all the diagnostic tests and rated all the mental effort, but their results were not used for the adaptation. After the participants finished the training session, a mental effort scale for the whole training session appeared. This mental effort rating was identical to the mental effort rating used before except for the wording. This one point assessment asked the participants to rate the mental effort invested during the entire training session.

![Figure 8. Structures of the stages during the training session](image)

**Final diagnostic tests**

The final diagnostic tests had exactly the same underlying structure as the initial diagnostic tests, but they were different problems containing dissimilar surface characteristics. Similar to the initial diagnostic tests, the final diagnostic tests contained three problems, one from each task class, and each problem has three sub-steps followed by a rating of subjective mental effort.

**Procedures**

This experiment was conducted as an assignment for one week. They individually participated in this experiment whenever and wherever they have an Internet access and computer. They earned 10% of the course credit on the basis of accomplishment of the whole learning package. The URL of the e-learning environment was provided to the participants by one instructor in class with specific instructions. Although there was no time limit controlled by the system, the participants were instructed to complete the whole program in two hours. This instruction was given in order to minimize any possible interruption in the middle of their learning. They were also instructed to try this
program only once, not to share their experience with other students and not to earn any assistance from other books or through the Internet. They were also informed of two types of programs, a text version and a text and audio version. They were instructed that as the system tailored the following questions according to their performance, if they did not know an answer, they had to select the “Don’t know” option. This instruction was given to minimize random selection of answers. All the participants’ responses were automatically recorded into a database created for this experiment. If participants were chosen for the bimodal WOE conditions, when they were on an instruction page for the training session, they could see a speaker icon and words saying, “There are audio files, so please prepare earphones.”

Research design and data analyses

The design of this study is a 2 by 2 factorial design. The two factors are the instruction methods (adaptive and fixed instructions) and the modality of WOE (bimodal and unimodal WOE). The dependent measures under the analyses were the pretest results, knowledge acquisition, mental effort, and instructional efficiency. Knowledge acquisition was computed by subtracting the initial diagnostic test scores from the final diagnostic test scores for each participant. The instructional efficiency was calculated as the ratios of knowledge acquisition to ratings of mental effort for the whole training session. PASW statistics 18 were used to perform ANOVA and ANCOVA. One-way ANOVA was conducted to compare group differences in the pretest results. The pretest mean scores of the four groups were not significantly different $[F(3, 92) = 0.60, p = .62]$. All dependent variables were initially analyzed with a two-way, between-subjects ANCOVA, taking the pretest results as a covariate. According to the results, the covariate revealed no significant relation to the dependent variables of knowledge acquisition, instructional efficiency, and instruction time, so the results from ANOVA were reported for these dependent variables. On the other hand, as the pretest results revealed significant relation to the dependent variables, the mental effort for the whole training session and during the final diagnostics tests, the results from ANCOVA were reported for those dependent variables.

Results and discussion

Is adaptive instruction more efficient than fixed instruction?

The adaptive instruction groups showed significantly higher instructional efficiency than the fixed instruction groups $[F(1, 92) = 4.68, p = .03, \eta_p^2 = .048]$ (see Table 2). In addition, there were significant differences in mental efforts for the whole training session and during the final diagnostic tests between the groups $[F(1, 91) = 6.53, p = .01, \eta_p^2 = .067, F(1, 91) = 7.77, p = .01, \eta_p^2 = .079]$. However, there was no significant difference in knowledge acquisition $[F(1, 92) = 0.34, p = .56]$. Learner expertise is associated not only with higher levels of performance but also with lower levels of cognitive load. As experts’ available knowledge structures in long-term memory could significantly reduce working memory demands, individuals with more expertise are able to attain equal or higher levels of performance with less investment of mental effort (Sweller, J., Ayres, P., & Kalyuga, S, 2011). Thus, it is likely that the level of expertise of the adaptive instruction groups in the domain of C programming language was higher after training than their counterpart. The significant difference in the mental effort during the final diagnostic tests supports this assumption as well. It implies that the learners who managed to acquire more knowledge during the training as a result of a more efficient instructional format experienced less cognitive load in test situations than learners who received a less efficient instructional format. Therefore, as expected, the effort to optimize the cognitive load according to the individual trajectory of cognitive skill acquisition successfully controlled cognitive load, thereby resulting in schema construction and automation. In addition, the rapid verification method coupled with a subjective mental effort rating seemed to probe the levels of learner expertise instantly and adequately in adaptive e-learning environments.

This result is consistent with the current literature on adaptive instructions within CLT. The study of Kalyuga and Sweller (2005) showed that the learner-adapted group to learner expertise showed significantly higher efficiency gain (Cohen’s $d$ effect size, 0.69) and significantly higher knowledge gain (Cohen’s $d$ effect size, 0.55). In addition, in Kalyuga’s (2006a) study, the learner-adapted group to learner expertise demonstrated a significantly lower mental effort and higher instructional efficiency than the non-adapted group, but there were no statistically significant
differences in knowledge acquisition. He argued that this lack of statistically significant differences between the experimental groups in knowledge acquisition could be due to a single training session. Likewise, in this study, the average amount of time the students spent in the training session was 24.4 min. More prolonged instructional events than a single training session may amplify knowledge acquisition of learners. From a slightly different perspective, Camp, Paas, Rikers, and van Merriënboer (2001) and Salden, Paas, Broers, & van Merriënboer, (2004) investigated adaptive instructions. They examined the assumption that the dynamic problem selection through efficiency measures would lead to more efficient training and better transfer than non-dynamic problem selection and dynamic problem selection based on performance or mental effort alone. These studies were primarily concerned with gradual increases in task difficulty rather than determining levels of guidance based on efficiency measures. However, their results also showed that the dynamic problem selection led to more efficient training than non-dynamic problem selection. Overall, the finding of this study provided the evidence that efficiency-based adaptation can be successfully and efficiently used to individualize instructional procedures in the domain of C programming language.

<table>
<thead>
<tr>
<th>Table 2. Effects of instruction methods on the dependent variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge acquisition</td>
</tr>
<tr>
<td>-----------------------</td>
</tr>
<tr>
<td>M (TS)</td>
</tr>
<tr>
<td>Adaptive</td>
</tr>
<tr>
<td>Fixed</td>
</tr>
</tbody>
</table>

Note. TS: Mental effort for the whole training session; FD: Mental effort during the final diagnostic tests

Is the bimodal process-oriented WOE more efficient than the unimodal process-oriented WOE?

There was no significant difference on knowledge acquisition between the bimodal and the unimodal WOE groups \( F(1, 92) = 0.34, p = .56 \). There were no significant differences in both mental efforts for the whole training session and during the final diagnostic tests \( F(1, 91) = 0.50, p = .48, F(1, 91) = 3.88, p = .05, \eta^2_p = .041 \) and in instructional efficiency, either \( F(1, 92) = 0.38, p = .54 \). Although no statistical support was found, as seen Table 3, the bimodal WOE groups acquired more knowledge, invested lower mental effort and showed higher instructional efficiency. In addition, the main effect of the modality of WOE on mental effort during the final diagnostic tests almost reached a significant level. This almost significant difference seems to imply that the learners who experienced less cognitive load and thereby managed to acquire more knowledge during the training experienced less cognitive load in test situations. Thus, it is fair to say that the bimodal WOE is more efficient than the unimodal WOE.

This finding is in line with the literature demonstrating the modality effect (Kalyuga, Chandler, & Sweller, 2000; Moreno & Mayer, 2002; Mousavi et al, 1995). In contrast, Tabbers et al. (2004) found that college students in the visual-only condition performed better than students in the audio condition on both retention and transfer tests and their interpretation for this reverse modality effect was that ecological classroom setting, longer studying time (more than an hour) and more procedure subject matter might have caused a reverse modality effect. In addition, they argued that a bimodal presentation was only advantageous when the instruction was system-paced whereas the visual-only instructions were the preferred format if the learners were in control. However, this study was not conducted in a laboratory setting and the participants spent a relatively longer time during a training session. The contents had both non-recurrent and recurrent components (procedural information) (van Merriënboer, Kirschner, & Kester, 2003) and the e-learning environment used was learner-paced. In spite of these similar experimental conditions, this study showed the modality effect.

<table>
<thead>
<tr>
<th>Table 3. Effects of modality of WOE on the dependent variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge acquisition</td>
</tr>
<tr>
<td>-----------------------</td>
</tr>
<tr>
<td>M (TS)</td>
</tr>
<tr>
<td>Bimodal</td>
</tr>
<tr>
<td>Unimodal</td>
</tr>
</tbody>
</table>

Note. TS: Mental effort for the whole training session; FD: Mental effort during the final diagnostic tests

268
Is there any interaction effect between instruction methods and modality of WOE?

There were no significant interaction effects between two factors on knowledge acquisition \([F (1, 92) = 0.28, p = .60] \), both mental efforts for the whole training session and during the final diagnostic tests \([F (1, 91) = 0.13, p = .72] \), and instructional efficiency \([F (1, 92) = 0.05, p = .83] \). As seen in Table 4, the bimodal process-oriented WOE group in the adaptive instruction showed the highest knowledge acquisition, lowest mental effort, and highest instructional efficiency. On the contrary, the unimodal WOE group in the fixed instruction showed lowest knowledge acquisition, highest mental effort, and lowest instructional efficiency. Thus, as expected, the bimodal process-oriented WOE provided necessary information for understanding in a cognitively efficient way, and as a result, utilization of the bimodal process-oriented WOE further improved the efficiency of the adaptive instruction.

<table>
<thead>
<tr>
<th></th>
<th>Knowledge acquisition</th>
<th>Mental effort</th>
<th>Instructional efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (TS)</td>
<td>SD (TS)</td>
<td>M (FD)</td>
</tr>
<tr>
<td>adaptive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bi</td>
<td>4.25</td>
<td>3.17</td>
<td>4.51</td>
</tr>
<tr>
<td>Uni</td>
<td>3.38</td>
<td>3.61</td>
<td>4.60</td>
</tr>
<tr>
<td>fixed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bi</td>
<td>3.38</td>
<td>4.15</td>
<td>5.47</td>
</tr>
<tr>
<td>Uni</td>
<td>3.33</td>
<td>4.38</td>
<td>6.05</td>
</tr>
</tbody>
</table>

Table 4. Interaction effect of instruction methods and modality of WOE on the dependent variables

Note. TS: Mental effort for the whole training session; FD: Mental effort during the final diagnostic tests; Bi: bimodal WOE, Uni: unimodal WOE

Conclusions and limitations

This study added evidence that tailoring instruction to the level of learner expertise through efficiency measures can be successfully used in e-learning environments in the domain of C programming language with college students. The adaptive instruction efficiently optimized cognitive load to the level of learner expertise and the efficiency measures based on the rapid verification method and its associated level of cognitive load successfully probed levels of learner expertise. Furthermore, the bimodal process-oriented WOE successfully controlled cognitive load and improved instructional efficiency further even in a learner-paced adaptive e-learning environment. These findings yield some important implications for future study. First, further research in the condition under which the modality effects occurs is necessary to uncover more sophisticated design principles in e-learning environments. Second, this study focused on knowledge acquisition (near transfer), but left out a possible effect on far transfer. The bimodal process-oriented WOE may yield effects on efficiency in far-transfer tasks. The issue of near and far transfer should be addressed in forthcoming studies. Third, the effects of the adaptive instruction with the bimodal WOE should be examined over a long extended period. The average amount of time for the training session was 24.4 min in this study. This length seemed too short to cause changes in learner knowledge structure. A prolonged practice may yield effects on knowledge acquisition. Finally, complex problems usually include both recurrent and non-recurrent components and it is not easy to make verification questions for non-recurrent components to probe learners’ knowledge structure activated in their long-term working memory. More research on solid diagnostic tests for non-recurrent components, at the same time appropriate for e-learning environments, is therefore necessary.

As a limitation of this study, the reliability estimates for both initial and final diagnostic tests did not prove to be sufficiently high. Each diagnostic test consisted of 9 questions and the relatively low reliability estimate was probably due to this very small number of test items. In addition, this experiment was conducted as an assignment. Specific instructions were given by the instructor to prevent any possible interruptions during their participation. However, interruptions happening in the middle of participations that could influences experimental results either positively or negatively could not be totally controlled for. The findings of this study are particularly important as excessive cognitive load prohibits a successful learning experience. Therefore, despite the limitations, insight gained through this study may inform instructional designers seeking to enhance their understanding of efficient instructional design within the CLT framework.
References


ABSTRACT

Open-ended learning environments (OELEs) are learner-centered, and they provide students with opportunities to take part in authentic and complex problem-solving tasks. These experiences may support deeper learning and the development of strategies that support future learning. However, many students struggle to succeed in such complex learning endeavors. Without proper adaptive scaffolding, these students often use system tools incorrectly and adopt suboptimal learning strategies. Developing adaptive scaffolds for students in OELEs poses significant challenges, and relatively few OELEs provide students with such support. This paper develops a model-based approach to interpreting and evaluating the actions students take as they learn in an OELE using a model of the cognitive and metacognitive processes that are important for completing the complex learning tasks. The model provides a means for classifying and assessing students’ learning behaviors as they work on the system, and it allows the system to identify opportunities to offer adaptive scaffolds to students. An evaluation of the analysis technique is presented in the context of Betty’s Brain, an OELE designed to help middle school students learn about science content.

Keywords
Open-ended learning environment, Model-based assessment, Adaptive scaffolding, Performance metrics

Introduction

Technological advances have provided researchers with affordances for designing computer-based learning environments that provide students with opportunities to take part in authentic, complex problem-solving tasks. These environments, generally called open-ended learning environments (OELEs) (Clarebout & Elen, 2008; Land & Hannafin, 1996; Land, 2000; Land, Hannafin, & Oliver, 2012), are learner-centered; they provide students with a learning context and a set of tools for exploring, hypothesizing, and building solutions to problems (Bransford, Cocking, & Brown, 2000). Examples include hypermedia learning environments (e.g., Azevedo, et al., 2012; Brush & Saye, 2001), modeling and simulation environments (e.g., Leelawong & Biswas, 2008; Sengupta, et al., 2013; van Joolingen, de Jong, Lazonder, Savenbergh, & Manlove, 2005), and narrative-centered learning environments (e.g., McQuiggan, Rowe, Lee, & Lester, 2008).

By the very nature of the choices they provide for learning and problem solving, OELEs are characterized by the opportunities they afford students to exercise metacognition (Clarebout & Elen, 2008; Land, 2000), which has been broadly defined as thinking about one’s own thinking. It involves two synergistic components: metacognitive knowledge and metacognitive regulation (Flavell, Miller, & Miller, 2002; Schraw, Crippen, & Hartley, 2006; Winne, 2001; Zohar & Dori, 2012). Together, these describe a person’s ability to explicitly set goals, establish plans for achieving goals using available resources, monitor progress toward achieving goals, and use the evaluation of progress to further regulate and improve one’s effectiveness. While OELEs may vary in the particular sets of tools they provide, they often include tools for: (i) seeking and acquiring information, (ii) applying information to a problem-solving context, and (iii) assessing the quality of the constructed solution (Lajoie & Azevedo, 2006; Moreno & Mayer, 2007).

OELEs place high cognitive demands on learners (Land, 2000). To be successful, learners must understand how to execute: (i) cognitive processes for accessing and interpreting information, constructing problem solutions, and assessing constructed solutions; and (ii) metacognitive processes for coordinating the use of cognitive processes and reflecting on the outcome of solution assessments. However, research has shown that students often lack regulatory processes necessary for achieving success (Hacker & Dunloskey, 2009; Zimmerman & Schunk, 2011). Without adaptive scaffolds, these learners typically use tools incorrectly and adopt suboptimal learning strategies (Mayer, 2004; Land, 2000). In this article, we define adaptive scaffolds in OELEs as actions taken by the learning
Developing adaptive scaffolds for students in these complex learning environments is a difficult task for system designers (Azevedo & Hadwin, 2005; Azevedo & Jacobson, 2008); it requires developing systematic analysis techniques for diagnosing learners’ needs as they relate to one or more cognitive and metacognitive processes. In OLEs, such diagnoses involve identifying and assessing learners’ cognitive skill proficiency, interpreting their action sequences in terms of learning strategies, and evaluating their success in accomplishing their current tasks. The open-ended nature of OLEs further exacerbates the problem; since the environments are learner-centered, they typically do not restrict the approaches that learners take to solving their problems. Thus, interpreting and assessing students’ learning behaviors is inherently complex; they may simultaneously pursue, modify, and abandon any of a large number of possible approaches to completing their tasks.

While several OLEs have been developed and used with learners, relatively few perform systematic interpretations of learners’ approaches to performing their tasks in order to provide adaptive support. Instead, these systems include non-adaptive scaffolded tools (e.g., lists of sub-goals or guiding questions) designed to provide support for learners who choose to use them (Puntambekar & Hübscher, 2005). In this paper, we discuss our recent work in developing and evaluating a novel model-based approach for measuring and assessing the actions students take as they learn with an OLE. The approach utilizes a model of the cognitive and metacognitive processes important for completing open-ended learning tasks in an effective manner. The model, then, provides a mechanism for the system to interpret students’ actions and behavior patterns in terms of these cognitive and metacognitive processes (e.g., Segedy, Kinnebrew, & Biswas, 2011), and it allows the system to identify opportunities to offer adaptive scaffolds to students. We illustrate our approach with Betty’s Brain (Leelawong & Biswas, 2008; Segedy, Kinnebrew, & Biswas, 2013), an OLE designed to help middle school students learn about science.

The remainder of this paper presents Betty’s Brain in more detail, including a description of the learning task and the cognitive and metacognitive model that characterizes student activities in such environments. We then use the model to derive a set of assessment metrics and apply those metrics post-hoc to data collected from a recent classroom study with Betty’s Brain. The goal is to demonstrate the utility of the model-based approach in interpreting students’ learning behaviors. In future work, we will incorporate our approach into the Betty’s Brain system to identify opportunities for providing adaptive scaffolds to learners as they work in the system. Finally, we discuss the implications of our results for the design of feedback and adaptive scaffolding for OLEs.

Overview of Betty’s Brain

Betty’s Brain, shown in Figure 1, presents the task of teaching a virtual agent, Betty, about a science phenomenon (e.g., thermoregulation in mammals) by constructing a causal map that represents that phenomenon as a set of entities connected by directed links that represent causal relations. Once taught, Betty can use the map to answer causal questions and explain those answers by reasoning through chains of links (Leelawong & Biswas, 2008). The goal for students using Betty’s Brain is to construct a causal map that matches a hidden, expert model of the domain.

As an OLE, Betty’s Brain includes tools for acquiring information, applying that information to a problem-solving context, and assessing the quality of the constructed solution. Students acquire domain knowledge by reading a set of hypertext resources that includes descriptions of scientific processes (e.g., shivering) and information pertaining to each concept that appears in the expert map (e.g., friction). As students read, they need to identify causal relations, such as “skeletal muscle contractions create friction in the body.” Students can then apply the learned information by adding the two entities to the causal map and creating the causal link between them (which “teaches” the information to Betty). In Betty’s Brain, learners are provided with the list of concepts and link definitions are limited to the qualitative options of increase (+) and decrease (-). Students can also add textual descriptions to each link.

Learners can assess their causal map by asking Betty to answer questions (using the pop-up window shown in Figure 1) and explain her answers. To answer questions, Betty uses qualitative reasoning methods that operate through chains of links from the source concept to the target concept (Leelawong & Biswas, 2008). When Betty explains her answers, she illustrates her reasoning by simultaneously explaining her thinking (e.g., the question said that the
hypothalamus response increases. This causes skin contraction to increase. The increase in skin contraction causes... and highlighting concepts and links on the map as she mentions them.

Figure 1. Betty’s Brain system with query window

After Betty answers a question, learners can ask Mr. Davis, another pedagogical agent that serves as the student’s mentor, to evaluate her answer. If Betty’s answer and explanation match the expert model (i.e., in answering the question, both maps utilize the same set of causal links), then Betty’s answer is correct. Note that a link’s textual description is not considered during this comparison. Learners can also have Betty take quizzes (i.e., sets of questions). Quiz questions are selected dynamically by comparing Betty’s current causal map to the expert map. Since the quiz is designed to reflect the current state of Betty’s map, a set of questions is chosen (in proportion to the completeness of the map) for which Betty will generate correct answers. The rest of her quiz answers are incorrect, and they are chosen to direct the student’s attention to parts of the map with missing or incorrect links. When Betty is unable to answer a question correctly, the students can use that information to discover and correct her misunderstandings. Similarly, when Betty answers a quiz question correctly, students know that the links that she used to answer that question are also correct. To help students in keeping track of which links are definitely correct, the system allows students to annotate causal links as being correct.

A model-based approach for assessing learning behaviors

As mentioned previously, properly supporting students in OELEs requires the ability to interpret their actions and behaviors. To accomplish this task systematically, we propose a novel model-based approach for measuring and evaluating the actions students take as they learn with an OELE. The approach utilizes a task model, developed by the authors and illustrated in Figure 2, of the cognitive and metacognitive tasks important for success in learning with OELEs (as discussed by Land, 2000 and Land & Hannafin, 1996, among others). The model specifies: (i) metacognitive tasks that learners need to engage in, (ii) cognitive processes that those tasks rely on, and (iii) tools in Betty’s Brain through which learners can enact their metacognitive tasks and cognitive processes. The metacognitive tasks in the model are open-ended and general; they apply to all OELEs, and learners need to employ metacognitive
regulation in order to accomplish them. In contrast, cognitive processes are procedural skills that learners need to employ properly while executing their plans, and they are specific to the OELE under study. The directed links in the model represent dependency relations. For example, solution assessment depends on learners’ abilities to infer which components of the causal map are correct and which are incorrect, which in turn relies on learners’ ability to interpret question grades, connect a question to the causal links that were used to answer the question, and evaluate Betty’s understanding. Performing these cognitive processes requires using the system tools that allow learners to: (i) ask Betty to take quizzes and explain answers, and (ii) ask Mr. Davis to evaluate the correctness of Betty’s most recent answer to a question.

The task model defines three broad classes of metacognitive tasks that learners need to engage in: (i) information seeking and acquisition, (ii) solution construction, and (iii) solution assessment. Each class of tasks involves a more specific set of metacognitive tasks, and each specific task could be accomplished by applying any of a number of metacognitive strategies. In terms of Bloom’s revised taxonomy (Krathwohl, 2002), these processes involve remembering and understanding information, analyzing it to identify information that could be useful in constructing a solution, applying it to the creation of a solution, and then submitting the solution to an automated assessment. Learners must then understand and analyze the assessment results in order to apply that additional information to keeping track of correct solution components and taking steps to refine incorrect solution components. As seen in the model, information seeking and acquisition tasks depend on one’s ability to identify and analyze the relevance of important information. Solution construction tasks depend on one’s ability to apply information gained from both the information seeking tools and the solution assessment results to creating and refining the causal map. Finally, solution assessment tasks depend on the learner’s ability to understand and interpret the results of solution assessments as actionable information that can be used to record progress and refine the current solution.

The task model provides a means for the system to interpret a student’s actions in terms of the tasks and processes defined by the model (Segedy, Kinnebrew, & Biswas, 2011; Kinnebrew & Biswas, 2012). As seen in the model, each tool in Betty’s Brain corresponds to multiple cognitive processes and metacognitive tasks. When a student uses one of these tools, their action is classified as being related to its associated processes and tasks. When a student accesses the resources, for example, the system interprets the action as being related to information seeking/acquisition, identifying causal relations in text, and correctly interpreting those causal relations. Similarly, when a student asks Betty to take a quiz, the action is interpreted as being related to solution assessment and evaluating Betty’s understanding.

Once classified, actions are further analyzed to assess aspects of students’ use of the cognitive and metacognitive processes described by the model. To assess cognitive processes, our approach judges each action taken on the system in terms of its effectiveness. Actions in an OELE are considered effective if they move the learner closer to their task goal, and effectiveness means something different for each class of metacognitive tasks. Effective information seeking, for example, helps students identify and acquire understanding of the specific domain content necessary for constructing an appropriate solution. Effective solution construction improves the overall quality of the solution in progress, and effective solution assessment produces and records information about the correctness and completion of the current solution. Some activities, particularly those related to information seeking, cannot always be automatically assigned an effectiveness score. However, through conversational feedback (Segedy, et al., 2013), the system could ask students to explain what they have learned from their activities, and the students’ responses can serve to inform the effectiveness metrics.

By combining interpretations from multiple actions over time, the system can infer more comprehensive aspects of student behavior related to their metacognitive activities. The analysis performed in this paper uses information about how learners’ actions can possibly cohere within logical plans to infer one aspect of students’ metacognitive planning: action support. In general, action support refers to whether or not an action taken by a learner was supported by previous actions. For example, information seeking actions (e.g., reading about a causal relationship) may provide support for future solution construction actions (e.g., adding that causal relationship to the map). Similarly, solution construction actions can be supported by information produced during solution assessment. This latter scenario may occur in Betty’s Brain when a student deletes a causal link from their map after Betty answers a question incorrectly on a quiz.

Together, classifying actions according to their related metacognitive tasks, cognitive processes, effectiveness, and action support allows the system to make decisions about when and how to scaffold learners as they work. If a
learner repeatedly performs ineffective actions related to a particular cognitive process, then the system can provide support targeted toward helping the learner improve their understanding of that process and how to use it more effectively. Similarly, if a student repeatedly performs unsupported actions related to a particular metacognitive task, then the system can scaffold students’ understanding of strategies for coordinating their use of the available tools in order to achieve a desired outcome.

**Post-hoc analysis using model-based assessments**

To illustrate the utility of the model based assessment approach, this section presents a post-hoc analysis of data from a recent classroom study conducted using Betty’s Brain. The goal is to use the model based metrics defined previously to examine how learners approached their learning tasks in terms of their understanding of cognitive processes and metacognitive tasks.

**Participants**

![Cognitive and metacognitive tasks model for OELEs](image1.png)

*Figure 2. Cognitive and metacognitive tasks model for OELEs*

![The Thermoregulation Expert Map](image2.png)

*Figure 3. The Thermoregulation Expert Map*
Fifty eighth grade students from three middle Tennessee science classrooms, taught by the same teacher, participated in the study. Because use of Betty’s Brain relies on students’ ability to independently read and understand the resources, the system is not suited to students with limited English proficiency or cognitive-behavioral problems. Therefore, while all students were encouraged to participate, data from ESL and special education students were not analyzed. Similarly, we excluded data from students who missed more than two class periods of work on the system. The final sample included data for forty students.

**Topic unit and text resources**

Students used Betty’s Brain to learn about mammalian thermoregulation in cold temperatures. The expert map (Figure 3) contained 13 concepts and 15 links representing cold detection (cold temperatures, heat loss, body temperature, cold temperature, hypothalamus response) and three bodily responses to cold: skin contraction (skin contraction, raised skin hairs, warm air near skin, heat loss), vasoconstriction (blood vessel constriction, blood flow to skin, heat loss), and shivering (skeletal muscle contractions, friction, heat generation). The resources were organized into two introductory pages, one page covering cold detection, and three pages covering the three bodily responses to cold temperatures. Additionally, a dictionary section discussed some of the concepts, one per page. The text was 15 pages (1,981 words) with a Flesch-Kincaid reading grade level of 8.9.

**Model-based assessments**

Students were assessed by utilizing the model-based methodology to interpret their use of cognitive and metacognitive processes while working with Betty’s Brain. Cognitive process assessments focused on students’ use of solution construction (causal map edits) and solution assessment tools, the latter of which were further divided into tools for model assessment (quizzes, question evaluations, and explanations) and progress recording (annotating links as correct). These assessment metrics were chosen mainly because the log data collected in this study can be readily used to compute them.

For each set of tools associated with a particular class of metacognitive tasks, two metrics were calculated for each student: (i) *skill application*, and (ii) *effectiveness*. Skill application is defined as the average number of times a tool related to a metacognitive task was executed per minute. For example, when applying the skill application metric to solution construction, the measure was computed as the average number of causal map edits performed per minute on the system. Effectiveness was calculated as the proportion of total actions related to a particular cognitive process that moved the students closer to their task goal. For solution construction, effectiveness refers to the percentage of causal link additions, removals, and modifications (with respect to the total number of such edits) that improved the quality of Betty’s causal map, where causal map quality is measured as the number of correct links minus the number of incorrect links in the map. For model assessment, effectiveness refers to the percentage of assessment actions that generated specific information about the correctness of one or more causal links. This may happen in one of several ways. First, a quiz may contain a question that Betty uses only one causal link to answer. In this case, the question’s grade directly corresponds to the correctness of the causal link. Second, learners can ask Betty to explain her answer to a question that Mr. Davis has graded as correct (either via a question evaluation or a quiz). When Betty answers a question correctly, all of the links she used to answer that question are also correct. Thus, when students ask Betty to explain her answer to a correct question, they generate correctness information for each link that Betty mentions in her explanation. Finally, learners may produce correctness information for a link via a guess and check strategy in which they have Betty take a quiz, note her quiz score, make a single change to the causal map, and have Betty take a second quiz. If the second quiz is higher or lower than the first, then students infer that their recent causal map edit was correct or incorrect, respectively. Finally, for progress recording, effectiveness refers to the percentage of link annotations created that correctly describe the annotated causal link (i.e., effective link annotations only describe a link as being correct when it actually is).

To assess one aspect of students’ metacognitive regulation, we calculated action support for each causal map edit. An edit was considered “supported” if one of the following conditions held: (i) students had previously accessed resource pages that discuss the concepts connected by the manipulated causal link, or (ii) the edit action removed or modified a causal link that had been previously proven incorrect via questions, question evaluations, quizzes, and explanations. A further constraint was added: an action could only support another action if both actions occurred
within a ten minute window. This resulted in three scores for each student: information seeking support percentage refers to the percentage of causal map edits that were supported by condition (i); model assessment support percentage refers to the percentage of causal map edits that were supported by condition (ii); and total support percentage refers to the percentage of causal map edits that were supported by either conditions (i) or (ii).

In addition to assessing the group of students as a whole, an additional set of analyses were employed to compare students who were more and less successful in teaching Betty the correct causal model. The goal was to investigate whether or not success in the system was associated with higher levels of effectiveness and action support. Students were divided into three groups based on the quality of their causal maps at the end of the study. Students in the Low group taught Betty a map that achieved a score of 5 or below. Students in the Medium group taught Betty a map with a score of 6 to 10, and students in the High group taught Betty a map with a score of 11 to 15 (where 15 is the maximum possible score). These resulting Low, Medium, and High groups contained 18, 6, and 16 students, respectively. Because so few students fell in the Medium group, the analysis focused on comparing the skill use, effectiveness, and action support metrics between only the Low and High groups.

Procedure

The study proceeded as follows: during the first 45-minute class period, the classroom teacher introduced students to thermoregulation. During the next class period, students completed a pre-test that included questions on thermoregulation and causal reasoning. During the next two classes, the research team introduced students to the causal reasoning method used by the system and provided students with hands-on system training. Students then spent five class periods (approximately 150 minutes) using Betty’s Brain with minimal intervention from the teachers and researchers. At the conclusion of the five class periods, all students took a post-test that was identical to the pre-test (pre-post test results have been reported in Kinnebrew & Biswas, 2012).

Results

Results of the group level cognitive process assessments are shown in Table 1. Students regularly engaged in information structuring and model assessment activities, editing their causal maps once every 2.28 minutes and assessing their map once every 5.15 minutes. However, students rarely made explicit records of the results of their assessment activities; they performed progress recording actions an average of once every 83.33 minutes.

<table>
<thead>
<tr>
<th></th>
<th>Actions/Min</th>
<th>Effectiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solution Construction</td>
<td>0.439 (0.190)</td>
<td>0.525 (0.113)</td>
</tr>
<tr>
<td>Model Assessment</td>
<td>0.194 (0.126)</td>
<td>0.370 (0.218)</td>
</tr>
<tr>
<td>Progress Recording</td>
<td>0.012 (0.033)</td>
<td>0.161 (0.347)</td>
</tr>
</tbody>
</table>

Despite engaging regularly in information structuring and model assessment behaviors, the students were not particularly effective in these endeavors. For information structuring, an average of just over half of their causal map edits improved the quality of Betty’s map. This suggests that students may have struggled to understand the causal relations described in the resources; alternatively, they may have edited their maps without first consulting the resources.

<table>
<thead>
<tr>
<th></th>
<th>Actions/Min</th>
<th>Effectiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question Evaluation</td>
<td>0.013 (0.019)</td>
<td>0.127 (0.280)</td>
</tr>
<tr>
<td>Explanation</td>
<td>0.033 (0.039)</td>
<td>0.060 (0.192)</td>
</tr>
<tr>
<td>Quiz</td>
<td>0.148 (0.111)</td>
<td>0.436 (0.233)</td>
</tr>
</tbody>
</table>

Students were even less effective in assessing Betty’s understanding of the science domain. On average, just under two thirds of their model assessment actions produced no information about the correctness of one or more causal links. Moreover, these actions were largely limited to quizzes; Table 2 shows that students asked for a question
evaluation once every 76.92 minutes, asked Betty to explain an answer once every 30.30 minutes, and asked Betty to take a quiz once every 6.76 minutes. This is striking, as quizzes by themselves don’t often provide correctness information for a causal link. Rather, quiz questions need to be combined with Betty’s explanations, which connect graded quiz answers to the sets of causal links that were used to generate those answers. It is important to note, however, that in some cases, students may have been able to infer which causal links generated an answer without requiring an explanation from Betty.

Table 3. Means (and standard deviations) of metacognitive process metrics

<table>
<thead>
<tr>
<th>Metric</th>
<th>Mean</th>
<th>(Standard Deviation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information Seeking/Acquisition Support %</td>
<td>60.2%</td>
<td>(18.5%)</td>
</tr>
<tr>
<td>Model Assessment Support %</td>
<td>0.8%</td>
<td>(01.4%)</td>
</tr>
<tr>
<td>Total Support %</td>
<td>60.9%</td>
<td>(18.3%)</td>
</tr>
</tbody>
</table>

Results of the group level metacognitive process assessments are shown in Table 3. Overwhelmingly, when students’ causal map edits had support, that support came from their recent information seeking/acquisition actions. Just over 60% of their causal map edits were supported by recent information seeking actions, as compared to less than 1% of their causal map edits being directly supported by model assessment actions. Thus, while students regularly quizzed Betty in ways that produced information about correct and incorrect links, they rarely took advantage of that information by deleting incorrect links or annotating correct links. This suggests that students may not have understood how to use model assessment tools in order to determine which links in their causal map were correct and incorrect. They may have alternatively decided to read the resources in order to investigate incorrect quiz questions, and this reading may have supported future editing activities.

Together, these results suggest that the students in this study struggled to use Betty’s Brain. More often than not, their solution construction, model assessment, and progress recording actions did not help them move toward their task goal. Moreover, almost 40% of their map edits were not supported by recent information seeking/acquisition or model assessment actions. However, despite these limitations, several students were successful (or close to successful) in teaching Betty the correct causal model. To explore the differences between students who were successful and those who were not, the cognitive and metacognitive process assessments were calculated for the High and Low student groups identified earlier. Because students rarely engaged in progress recording activities, the analysis focused only on information structuring and model assessment.

The results of the comparative analysis (Figure 4) show that the High group students more often performed actions linked to information structuring. Additionally, the High group’s information structuring actions were more likely to be effective and have support from related activities. Independent samples t-tests conducted on these data showed a significant effect of group on actions per minute, \( t(32) = 2.67, p = .012 \), effectiveness, \( t(32) = 5.99, p < .001 \), and total support percentage \( t(32) = 2.57, p = .015 \). Figure 4 also shows that the High group performed slightly more model assessment actions than did the Low group; however, the Low group’s model assessment actions were more likely to produce information about the correctness of one or more links. Independent samples t-tests conducted on these data did not reveal a significant effect of group on actions per minute, \( t(32) = 0.98, p = \text{n.s.} \), or effectiveness, \( t(32) = 0.79, p = \text{n.s.} \).

![Figure 4. Means (and standard deviations) of process metrics for high and low groups](Image)
These results show that High group students’ superior performances were associated with a higher rate of causal map editing and a higher proportion of effective and supported edits. However, their success was not associated with effective model assessment activities, and their map edits were rarely supported by model assessment actions. One possible interpretation is that these students were more effective than students in the Low group in identifying causal links when reading the resources. However, the low effectiveness in map editing activities combined with the extremely low number of causal map edits supported by model assessment activities in both groups implies that even students who successfully completed the learning task may have benefited from adaptive scaffolds encouraging them to carefully and systematically utilize the results of quizzes, question evaluations, and explanations.

Discussion and conclusions

In this paper, we have presented a novel model-based approach for analyzing the actions students take as they learn with an OELE. The approach characterizes OELEs as environments in which learners must utilize cognitive and metacognitive processes as they seek out and acquire information, apply that information to the construction of a solution, and assess the quality of their solution. We incorporated this characterization into a task model, which we then used to motivate metrics of effectiveness and action support. Effectiveness indicates whether or not learners’ actions are advancing them toward their goal, and action support measures whether or not a particular action could have been motivated by information generated during a previous action. We then applied these metrics to perform a post-hoc analysis of data collected from Betty’s Brain.

The analysis found that while students frequently used system tools related to solution construction and model assessment, their use of these tools was often sub-optimal. A large proportion (47.5%) of their causal map edits decreased the quality of their maps, and an even larger proportion (63.0%) of their model assessment activities did not produce information related to the correctness of one or more causal links. Further, 39.1% of their edits were not supported by recent information seeking/acquisition or model-assessment activities. Even students who were able to successfully teach Betty the correct causal map performed a large proportion of incorrect map edits (38.2%), unsupported map edits (30.3%), and ineffective model assessment actions (78.9%). The results of this analysis reveal that the students using Betty’s Brain in this study might have benefited from adaptive scaffolds targeted toward helping students employ more effective strategies as they worked toward completing the learning task.

Importantly, the results of applying our model to analyze data from Betty’s Brain demonstrate the utility of the analytic framework presented in this paper. By using just a subset of possible metrics for assessing students’ use of cognitive and metacognitive processes, the analysis provided a detailed description of how learners in this study approached the open-ended learning task. Such information will be critical in improving the adaptive scaffolding in Betty’s Brain. The model-based analyses can be executed in real-time as learners are working in Betty’s Brain, and, once we incorporate them into the system, Mr. Davis will be able to leverage them in deciding how to scaffold learners.

There are a number of other OELEs that assess aspects of students’ metacognition. For example, MetaTutor (Azevedo et al., 2012) is a hypermedia learning environment in which learners navigate information sources in order to achieve a pre-determined overarching learning goal (e.g., learn everything you can about the human circulatory system). To accomplish their goal, learners must employ a variety of metacognitive strategies for setting sub-goals, seeking out relevant information, and evaluating sources of information as they are encountered. The system infers learners’ goals and strategies by prompting them to explicitly communicate those goals and strategies to the system as they work. For example, a learner might set a sub-goal of learning about the structure of the human heart and then select the strategy perform a goal-directed search from a menu of options called the Self-Regulated Learning Palette. As another example, recent work with Crystal Island (Sabourin, Shores, Mott, & Lester, 2012) periodically required students to indicate their affective state and a “status update.” A post-hoc analysis of these self-reports found that they were predictive of student success. Thus, the authors concluded, the information provided by these reports could be used to target scaffolding and feedback to the needs of learners.

Both of these techniques require continuous self-reporting from students during learning, which differs from the approach taken in this paper. However, incorporating self-report measures into the model-based approach could lead to the ability to derive more informative assessment metrics. If learners using Betty’s Brain were required to indicate their current sub-goal to the system, the system could use that additional information to select and apply a set of
assessment metrics related to accomplishing that particular sub-goal. For example, if a learner has indicated that she is re-reading to make sure that she correctly constructed a causal link, then an assessment metric may test her understanding of the material related to only that link.

Future work with model-based analysis in OELEs will develop a more comprehensive set of effectiveness and support metrics with which to evaluate learners as they use Betty’s Brain. Additionally, we will incorporate these metrics into Betty’s Brain in order to identify opportunities for providing adaptive scaffolds to learners based on their needs. Once the decision to support students in using a particular type of cognitive or metacognitive process has been made, the pedagogical agents in the system will use the assessment metrics to drive a discussion toward identifying exactly why the learner is having trouble. Ideally, the agents will be able to provide students with the scaffolding they need to gain a better understanding of the cognitive and metacognitive processes important for managing one’s own learning process. To that end, we have developed a set of tutorial activities for students to practice some skills important for success in teaching Betty, such as identifying causal relations from reading materials and identifying relevant link correctness information from quizzes (Segedy, Biswas, Blackstock, & Jenkins, 2013). Hopefully, this support will lead to an increase in students’ effectiveness in learning with Betty’s Brain.

Acknowledgements

This work has been supported by Institute of Educational Sciences CASL Grant #R305A120186 and the National Science Foundation’s IIS Award #0904387.

References


Learning Preferences and Motivation of Different Ability Students for Social-Competition or Self-Competition

Zhi-Hong Chen

Department of Information Communication, Yuan Ze University, Taiwan // zhchen@saturn.yzu.edu.tw

(Submitted February 5, 2013; Revised April 22, 2013; Accepted May 29, 2013)

ABSTRACT

Competitive learning is attracting ever increasing amounts of attention in the field of digital game-based learning. Different mechanisms for the promotion of competitive learning have been proposed in previous studies, including social-competition and self-competition mechanisms, but few have addressed student preferences as to the choice between social-competition and self-competition, especially considering students’ different levels of capabilities and their perception. Thus, this study investigates how students with different levels of capabilities choose and perceive learning models between social-competition and self-competition. The study was carried out using the mixed-model experimental design. Sample consisted of 54 elementary school students assigned into three ability-level groups with all groups experiencing both treatments of social- and self-competition digital game-based learning environments. The results indicate that low-ability students had lower test anxiety and greater preference for social-competition, whereas medium-ability and high-ability students showed higher test anxiety and a similar preference for social-competition and self-competition. Based on this result, competitive learning design framework should consider enjoyment aspect of social competition for low-ability students, and interactive and performance aspects for self- and social-competition for medium- and high-ability students.

Keywords

Games, Primary school, Interactive learning environments

Introduction

Competition is a powerful motivator for student behavior regularly applied in education research. Some studies have found competition to have positive effects on student learning (Cheng, Wu, Liao, & Chan, 2009; Yu & Liu, 2009; Yu, Han, & Chan, 2008; Ke, 2008a; 2008b). However, competition could also have negative effects on student confidence and learning development (Staples & Koomen, 2005; Mussweiler, 2003; Gilbert, Giesler, & Morris, 1995). This could be because competition involves social comparisons, where students are exposed to conflicting sets of comparative information and processes, including failures in the competition. In particular, most competitions are zero-sum activities, in which one student loses the competition while the other wins. If the loser attributes this failure to lack of ability, he/she might feel frustrated or helpless (Dweck, 2000; Weiner, 1986; 1985).

To prevent possible negative effects, previous studies have proposed some mechanisms. For example, an anonymous mechanism is used to diminish the negative impact resulting from a face-to-face competitive setting (Yu & Liu, 2009; Yu, Chang, Liu, & Chan, 2002) or a surrogate mechanism is used to foster learning effort through commanding virtual pets to compete against each other (Chen, Liao, Chien, & Chan, 2011; Chen et al., 2012). In addition to these social mechanisms, a self-competition mechanism is also proposed, where students compete against themselves rather than their learning peers. Since students within the self-competition setting are not subject to any level of social pressure, possible negative impacts on students might be diminished.

However, the current studies seem to lack a systematic examination of students’ learning preferences and motivation for the choices between social-competition and self-competition. This issue is significant because it could serve as a foundation to guide the development of competitive learning models. In this vein, this study investigates students’ choice preferences and motivation for social-competition and self-competition. In addition, since students’ ability level often plays a significant role in competition, we also take students’ ability level into account in this study. Specifically, we try to answer the research question: What are the learning preferences and motivation of different ability students for social-competition or self-competition? The rest of this paper is organized as follows. Section 2 presents related works about competitive learning. Section 3 describes an experiment, which investigates students’ preferences for social-competition or self-competition. In Section 4 we discuss the results of the experiment, which are
further used to develop a design framework for competitive learning. Finally, some conclusions and the limitations of this study are offered in Section 5.

Related work

Competition is regarded as useful for student learning because it can reinforce the goal structure of learning activities (Davis & Rimm, 1985), which, in turn, enhances students’ motivation and learning achievement (Ke, 2008a). Nevertheless, the use of competition can also have negative influences (Stapel & Koomen, 2005), such as the lack of a scheme for improvement (Chan et al., 1992) and a high degree of stress (Yu & Liu, 2009). This is because competition involves a comparison process, during which participants are compared with each other (Martens, 1976). Such comparisons can affect students’ confidence, attitudes, and belief in success (Mussweiler, 2003). In addition, most competition occurs under specific conditions where some students lose while others win. In such competition the loser may feel hurt as a result. Thus, there is a need to take the negative effects of competition into account.

To this end, a number of competitive mechanisms have been investigated, as illustrated in Table 1. The first three emphasize the individual model, whereas the latter three focus on the social model. Regarding the individual model, the common goal of these mechanisms is to help students be more aware of learning status and further take actions to improve through different approaches, such as the improving space (Chan & Lai, 1995), learning companion (Chan & Lai, 1995), and avatar (Chen, Chien, & Chan, 2011). These mechanisms are underpinned by the theory of self-regulated learning (Schunk, & Zimmerman, 1998): offering students suitable information to foster their strategy planning, progress monitoring, and outcome evaluation. In other words, the students’ self-generated thoughts, feeling, strategies, and behaviors might contribute to a better learning cycle, in which the students could play an active and dominant role in their learning.

Regarding the social model, the design focus has shifted from self-regulated learning to social interaction underpinned by the hypothesis of zone of proximal development (Vygotsky, 1978), a distance between what students can achieve by themselves and what they can achieve when provided with appropriate support through social interactions with peers. Thus, different social mechanisms are incorporated with competition to support student learning. For example, the Teams Games Tournaments (TGT) allows students to work together as a learning group to compete against other groups (Slavin, 1990; Ke, 2008a; Ke, 2008b). Other related mechanisms include the anonymous mechanism (Yu et al., 2008) and surrogate mechanism (Chen, Chow, Biswas, & Chan, 2012). The two mechanisms help students alleviate possible negative effects by hiding identities and virtual surrogates, respectively.

### Table 1. Some mechanisms for competitive learning

<table>
<thead>
<tr>
<th>Mechanisms</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual models</td>
<td></td>
</tr>
<tr>
<td>Improving space</td>
<td>Encourages students to prepare themselves better before the competition takes place so that they have more chances to win.</td>
</tr>
<tr>
<td>(Chan &amp; Lai, 1995)</td>
<td></td>
</tr>
<tr>
<td>Learning companion</td>
<td>Offers computer-simulated agents to help students improve their learning performance and win the competition.</td>
</tr>
<tr>
<td>(Chan &amp; Lai, 1995)</td>
<td></td>
</tr>
<tr>
<td>Avatar mechanism</td>
<td>Provides students with avatars to represent their learning performance in different units so that they can compete with themselves.</td>
</tr>
<tr>
<td>(Chen et al., 2011)</td>
<td></td>
</tr>
<tr>
<td>Social models</td>
<td></td>
</tr>
<tr>
<td>Anonymous mechanism</td>
<td>Hides students’ identities as a protective mechanism for preventing negative effects when they lose in the competition.</td>
</tr>
<tr>
<td>(Yu et al., 2008)</td>
<td></td>
</tr>
<tr>
<td>Group mechanism</td>
<td>Shares the risk and responsibility between group members to prevent negative effects when students lose in the competition.</td>
</tr>
<tr>
<td>(Ke, 2008a; 2008b)</td>
<td></td>
</tr>
<tr>
<td>Surrogate mechanism</td>
<td>Offers virtual pets as mediators in competition for the shaping of positive attribution and belief in effort.</td>
</tr>
<tr>
<td>(Chen et al., 2012)</td>
<td></td>
</tr>
</tbody>
</table>

Furthermore, competition has a great impact on students’ performance, confidence, or even attitudes. It has been revealed that students’ academic performance has a negative relationship with their test anxiety (Chapell et al., 2005), which might result from a higher-pressure setting. In other words, students with higher pressure or anxiety might lead to lower performance. By contrast, some studies demonstrated that the relationship between anxiety and performance is an inverted U-shaped curve (Yerkes & Dodson, 1908), implying that appropriate arousal of anxiety could contribute to students’ learning. Thus, when competitive mechanisms are applied to educational settings,
different students might perceive competition as different levels of pressure, which might further influence their preferences and choices.

In such a vein, a significant research question arises while reviewing those related studies: What are students’ preferences and motivation when using these competition-based models? This research question is significant because students’ choices and preferences to use learning systems is a prerequisite for effective learning outcomes. Nevertheless, this question has seldom been addressed in past studies, especially within computer-supported competitive models. If we could have a better understanding of different students’ choice preferences and motivation for the two models, it would be helpful to design suitable competitive models that take individual differences into account. Thus, this study focuses on addressing students’ preferences and motivation for individual or social competition-based models.

Methodology

To address the research question, a mixed-model experimental design was conducted in this study. More specifically, there are two independent variables: the levels of ability and competition models. Each student with different levels of ability (i.e., between-subject variable) experienced the two competition models (i.e., within-subject variable) and gave feedback to their choices. In particular, the avatar mechanism and surrogate mechanism serve as two examples of self-competition and social-competition models, respectively.

Participants

In total 54 elementary school students (aged an average of 11 years) in Taiwan participated in the experiment. To further understand whether students’ learning ability would influence their preferences, these students were further categorized into three different levels of ability based on their examination scores in school in the subject of language learning, ranging is from 0 to 100. Specifically, the method of categorization into the three levels (i.e., low, medium, or high ability) contains two steps: (1) the total number of students is divided by three groups, 54/3=18 (2) If the scores of the students with two different levels are the same, these students are re-assigned to an upper group. Finally, the numbers of students in the three groups are 16, 17, and 21, respectively. The details are described in Table 2.

<table>
<thead>
<tr>
<th>Student ability</th>
<th>Number of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level L (low-ability)</td>
<td>16</td>
</tr>
<tr>
<td>Level M (medium-ability)</td>
<td>17</td>
</tr>
<tr>
<td>Level H (high-ability)</td>
<td>21</td>
</tr>
</tbody>
</table>

Table 2. Students in each ability level

Instruments

Two system instruments

Two learning systems were used in the experiment, in which the subject domain (Chinese idioms) was the same in both versions. The major difference between the two systems lies in the opponents that students compete against. Specifically, as illustrated in Table 3, students who used the self-competition system learn Chinese idioms (Table 3b) and their learning status is shown through avatars (Table 3a). These avatars represent the learning outcomes for different units as the avatars compete against each other (Table 3c). Similarly, the learning status of students who used the social-competition version to learn Chinese idioms (Table 3p) is also shown in their avatars (Table 3q). However, a student can choose one of the other students as an opponent (Table 3r) so that his or her avatar can compete against that of the other student (Table 3s).
<table>
<thead>
<tr>
<th>Virtual character</th>
<th>Self-competition</th>
<th>Social-competition</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) An avatar is used to represent the student’s current learning status</td>
<td>(p) An avatar is used to represent the students’ current learning status</td>
<td></td>
</tr>
<tr>
<td>Learning material</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(b) Chinese idiom learning activities</td>
<td>(q) Chinese idiom learning activities</td>
<td></td>
</tr>
<tr>
<td>Competitive model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(c) The avatar competes against student’s own avatar for previous units</td>
<td>(r) The student chooses one of the other students as an opponent</td>
<td></td>
</tr>
<tr>
<td>(s) The avatar competes against another student’s avatar</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Motivated Strategies for Learning Questionnaire (MSLQ)**

A motivational questionnaire adapted from the Motivated Strategies for Learning Questionnaire (Pintrich, Smith, Garcia, & McKeachie, 1991) with a 7-point Likert choices was used to collect students’ motivational feedback with choices ranging from “not at all true for me” (point=1) to “very true for me” (point=7). This information provides explanation on the students’ choice preferences. The motivational scale of the MSLQ contains three aspects: value, expectancy, and affective. Regarding the value component, there were 4 items for intrinsic goal, 4 items for extrinsic goal, and 6 items for the task value aspect; regarding the expectancy component, there were 4 items for control of belief, and 8 items for the self-efficacy aspect; regarding the affective component, there were 5 items for the aspect of test anxiety. According to the work of Pintrich and his colleagues (1991), this questionnaire had a relatively high reliability (Cronbach’s alpha were .74, .62, .90, .68, .93, and .80 for the six aspects, respectively).
Preference Questionnaire (PQ)

To collect students’ preferences as to which competitive system they preferred and the reasons why they made this choice, a perception questionnaire (developed by the first author of this paper) was used. The questionnaire contained two items: “Which system version did you prefer to use?” (i.e., students are asked to choose one between self-competition and social-competition systems) and “For what reason did you prefer to use it?” (i.e., students are asked to write down theirs reasons for this open-ended question).

Procedures

To reduce the bias of treatment order, participants were divided into two groups, who used the different system versions in a different order, as shown in Figure 1. Each session lasted 40 minutes and was held once a week. The employed procedures are as follows: (1) before the experiment, participants were instructed how to use the systems. In addition, to increase the validity of their opinions, the participants were also told that they could freely use the system functions; (2) during the sessions, participants in the two groups used different systems. The Group A students first used the self-competition system for one session and then used the social-competition system for another session. In contrast, the Group B students first used the social-competition system and one week later switched to the self-competition system. (3) The MSLQ was immediately administrated at the end of each session to collect information on students’ motivation. In other words, the Group A students were asked to complete the MSLQ twice: at the end of self-competition session and social-competition session, respectively. Similarly, the Group B students were also asked to complete the MSLQ twice. (4) At the end of all sessions, participants were further asked to fill out the PQ. In this way, participants’ preferences for the two system versions were collected after they have experienced the use of both systems.

Data analysis

The data analysis comprises two parts. For the motivational scale, a repeated measures analysis of variances (ANOVA) was carried out, with two system versions (i.e., self-competition and social-competition) and three levels of ability (i.e., level L, M, and H) as independent variables and the motivational aspect as the dependent variable. All these analyses were conducted with the Statistical Package for the Social Science (SPSS Windows version 17).

For the preference questionnaire, Chi-square tests were conducted to validate the significant difference between students’ preference between two groups, with the two system versions as independent variable and the number of students as dependent variable. In addition, content analysis was used to categorize students’ reasons for preferring the chosen systems. Since no existing and suitable coding scheme for competition-based learning models could be found in the literature, this study uses a coding scheme based on the two previous studies: the reasons for loneliness and social use of the Internet (Morahan-Martin & Schumacher, 2003) and the sources of competence information (Horn, 2008; Horn & Amarose, 1998). This is because competition-based learning involves both technology use and theory of competence and motivation (Harter, 1978; 1981).

Specifically, the coding scheme is organized into three aspects: (1) the cognitive aspect on the level of the individual effort; (2) the social aspect related to the degree of interpersonal relationship; and (3) the affective aspect on the
game-related feelings. Table 4 illustrates the details of the coding scheme. Two graduate students were trained to use the scheme to independently code all responses to the preference questionnaire. A statement was deemed “codeable” if the two coders used the same coding scheme. If this was not the case, they discussed the choices until they came to an agreement. The Cohen’s kappa of the coder inter reliability was .79, representing good agreement beyond chance (De Wever, Schellens, Valcke, & Van Keer, 2006).

Table 4. The coding scheme used in this study

<table>
<thead>
<tr>
<th>Cognitive aspect</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>a1: self-awareness</td>
<td>Enhances the ability or knowledge to understand students’ learning status or performance better</td>
</tr>
<tr>
<td>a2: self-comparison</td>
<td>Emphasizes the comparative or competitive process, in which a student’s status at different stages is examined</td>
</tr>
<tr>
<td>a3: effort-exerting</td>
<td>Highlights the learning effort that the student makes</td>
</tr>
<tr>
<td>a4: self-improvement</td>
<td>Focuses on improving learning status better than before, or surpasses past performance</td>
</tr>
</tbody>
</table>

Social aspect

<table>
<thead>
<tr>
<th>Social aspect</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>b1: self-protection</td>
<td>Focuses on the lack of confidence to interact with others</td>
</tr>
<tr>
<td>b2: knowing-others</td>
<td>Enhances the ability or knowledge to understand other students’ learning status or performance better</td>
</tr>
<tr>
<td>b3: peer-comparison</td>
<td>Emphasizes the comparative or competitive process, in which the students’ statuses are examined to determine who is better</td>
</tr>
<tr>
<td>b4: peer-interaction</td>
<td>Enhances interactivity through being (or playing) with peers/friends, rather than competition/comparison</td>
</tr>
</tbody>
</table>

Affective aspect

<table>
<thead>
<tr>
<th>Affective aspect</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>c1: game-enjoyment</td>
<td>Highlights the feeling of fun and satisfaction resulting from the general features of a game or an activity</td>
</tr>
<tr>
<td>c2: social-enjoyment</td>
<td>Emphasizes the feeling of fun and satisfaction resulting from the interpersonal features or design of a game or an activity</td>
</tr>
<tr>
<td>c3: challenge</td>
<td>Focuses on the feeling of challenge in an activity</td>
</tr>
<tr>
<td>c4: easy-to-use</td>
<td>Focuses on the feeling of ease of system usability</td>
</tr>
</tbody>
</table>

Results and discussion

Motivational levels in using Digital Game-Based Learning

Table 5 illustrates the mean and SD of the scores on the motivational scale in terms of six aspects for all participants. The results of ANOVAs show that there was a statistically significant difference only for the aspect of test anxiety, whereas there were no significant differences for other aspects. Furthermore, regarding the aspect of test anxiety, the results showed that there was no significant difference for the effect of different systems (F(1,51) = .134, p = .715), whereas there was a significant difference demonstrated for the effect of different ability levels (F(2,51) = 4.51, p < .05), in which the interaction effect of the two variables was not significant (F(2,51) = .488, p = .616), implying that participants with different knowledge levels perceived different levels of test anxiety. The result of a further Tukey HSD post-comparison revealed that the scores for Level M and H were significantly larger than those for Level L (p < .05). This implies that participants with high- and medium-ability had higher test anxiety than those with low-ability, regardless of whether in the self-competition or social-competition category.

Table 5. The results of the motivational scale for all participants

<table>
<thead>
<tr>
<th>Value component</th>
<th>Self-competition</th>
<th>Social-competition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Level L</td>
<td>Level M</td>
</tr>
<tr>
<td>Intrinsic goal</td>
<td>5.50(1.3)</td>
<td>5.91(1.0)</td>
</tr>
<tr>
<td>Extrinsic goal</td>
<td>5.14(1.4)</td>
<td>5.97(2.0)</td>
</tr>
<tr>
<td>Task value</td>
<td>5.40(1.3)</td>
<td>6.00(0.8)</td>
</tr>
</tbody>
</table>

Expectancy component

288
Preferences for self- and social-competition

Table 6 illustrates the number of students and their reasons for the selecting the chosen systems. For the low-ability (i.e., level L) students, the ratio of self-competition to social-competition was one to three. The results of the Chi-square test indicate that their preference was significantly different ($\chi^2 = 4.00$, df = 1, $p < .05$), implying that those students preferred social-competition. It should be noted that the major reason given for why students preferred either self-competition or social-competition was the same: they found the system more fun or exciting (N = 2 and N = 7, respectively). In other words, it was the *enjoyment* aspect of the competitive activities that dominated the low-ability students to choose these systems.

A possible interpretation for this was a “self-protective” viewpoint, in which these students utilized escape or avoidance behavior to reduce their stress and anxiety (Zeidner, 1998). In other words, the low-ability students might perceive the obvious interaction of the social-competition system to be more like a game, rather than a serious test. Thus, most students chose the social-competition system. If they perceived that what they were doing was a part of game-playing, rather than a serious assessment process, they would feel more relaxed, leading to lower test anxiety. This could also explain why low-ability students expressed lower test anxiety in the motivational scale.

<table>
<thead>
<tr>
<th>Levels</th>
<th>Preferred models</th>
<th>N</th>
<th>Reasons (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level L</td>
<td>Self-competition</td>
<td>4</td>
<td>c1 game-enjoyment (2) a2 self-comparison (2)</td>
</tr>
<tr>
<td></td>
<td>Social-competition</td>
<td>12</td>
<td>c1 game-enjoyment (4) c2 social-enjoyment (3) b4 peer-interaction (2) b3 peer-comparison (1) c4 easy-to-use (1) a4 self-improvement (1)</td>
</tr>
<tr>
<td>Level M</td>
<td>Self-competition</td>
<td>6</td>
<td>a4 self-improvement (3) a2 self-comparison (1) a1 self-awareness (1) b1 self-protection (1)</td>
</tr>
<tr>
<td></td>
<td>Social-competition</td>
<td>11</td>
<td>c1 game-enjoyment (3) c2 social-enjoyment (2) b4 peer-interaction (2) b3 peer-comparison (2) b2 knowing-others (1) c3 challenge (1)</td>
</tr>
<tr>
<td>Level H</td>
<td>Self-competition</td>
<td>10</td>
<td>a4 self-improvement (3) a2 self-comparison (3) a3 effort-exerting (3) a1 self-awareness (1)</td>
</tr>
<tr>
<td></td>
<td>Social-competition</td>
<td>11</td>
<td>b3 peer-comparison (3) b2 knowing-others (3) b4 peer-interaction (2) c2 social-enjoyment (2) a1 self-awareness (1)</td>
</tr>
</tbody>
</table>

For the medium-ability (level M) students, the ratio of self-competition to social-competition was approximately one to two. The results of the Chi-square test indicated no significant difference in their preference ($\chi^2 = 1.47$, df = 1, $p < .289$).
suggesting similar preferences. Like the low-ability students, some preferred the social-competition system, but their reasons were different. The major reason for selecting self-competition was that this system could enhance self-improvement (N = 3), whereas the major reason for preferring social-competition was that it could support social interaction (i.e., game-enjoyment, N = 3; social-enjoyment, N = 2; peer-interaction, N = 2; peer-comparison, N = 2; knowing-others, N = 1). The result seemed to imply that the medium-ability students chose their systems based on the interactive aspect of the competitive activities—that is whether this competitive activity is individual or social? One possible explanation was that the medium-ability students who chose the self-competition had higher self-discipline and self-efficacy (Bandura & Jourden, 1991). In other words, the students with higher ability have higher self-efficacy and the students believed that they could well manage learning and control behavior on their own. The self-competition system offered them more information to understand their learning status, so they preferred this version. On the other hand, the students who chose the social-competition could be social-directed learners, who regarded social interaction as the key element for a competitive activity. Previous studies indicate that the majority of people socialize in a game activity (Zichermann & Cunningham, 2011; Bartle, 1996), which seems to be consistent with the results obtained in this study.

With the high-ability (i.e., level H) students, the ratio of self-competition to social-competition was almost the same. The result of the Chi-square test did not show a significant difference ($\chi^2 = .04, df = 1, p > .05$), implying that those students preferred both versions. The main reason for the choice of self-competition was that it could enhance self-improvement (N = 3) and self-comparison (N = 3), which are the same as for the medium-ability students described above. However, unlike the medium-ability students, the high-ability students paced more emphasis on effort-exerting (N = 3), suggesting that they seemed to be more self-disciplined. On the other hand, the main reason for the social-competition preference was still social interaction (i.e., peer-comparison, N = 3; knowing-others, N = 3; peer-interaction, N = 2; social-enjoyment, N = 2). Game-enjoyment, a significant reason given by the medium-ability students, was not found for the high-ability students. The implication is that game-playing is not as significant for them as for the low-ability and medium-ability students. Considering the aforementioned system preferences, it appears that the high-ability students tended to emphasize on the performance aspect of the competitive activities—that is whether this competitive activity is helpful to their learning performance or not.

A reasonable interpretation for these results is that the high-ability students seemed to regard competition as a serious activity, rather than a game-playing activity. Thus, students who highlighted self-improvement chose the self-competition system while others regarded an indirect (i.e., collecting public information about others) or direct form of interpersonal interaction (i.e., comparing or competing with others) as a significant part of a learning activity so they preferred the social-competition system. In other words, those students treated competition as an activity for evaluating their performance, which might explain why they had higher test anxiety.

**Suggestion for design framework**

Figure 2 illustrates the design framework based to the results of this study, which should contribute to the design of future competitive learning models.

![Figure 2. Design framework for competitive learning models](image)
Low-ability students seemed to regard the competition as a game-playing activity, in which the enjoyment aspects (i.e., fun or excitement of competitive activities) are crucial. In other words, these students were concerned with the enjoyment aspect of competitive activities, rather than learning performance. Thus, they also experienced lower test anxiety. For those students, there is a need to take the enjoyment element of competitive activities into account when designing self-competition systems in the future.

As for medium-ability students, they tended to emphasize the interactive aspect of competitive activities. Specifically, some preferred social-competition because it could offer them the opportunities to interact with others, whereas others liked self-competition because it was helpful for self-improvement. Nevertheless, compared with the low-ability students, these students experienced higher test anxiety, perhaps due to the fact that enjoyment was not their major concern. Instead, these students chose their competitive activities based on their preference for the interactive aspects. Consequently, how to, on the one hand, enhance the interactivity of learning activities and, on the other hand, reduce students’ test anxiety is a critical issue for designing competitive learning activities in the future.

Of the high-ability students, some chose self-competition because of its helpfulness to their learning performance while some chose social-competition because of its contribution to social interaction. The students had a high-level of test anxiety whether in self-competition or social-competition activities, suggesting that they regard competitive activities as serious assessments, rather than game-playing. In other words, it was more important to improve performance, rather than enjoy the activity or interaction. Consequently, how to provide those students with suitable information to help them reflect what they have learned, regulate their efforts, and improve their learning status either for self-competition or social-competition is an emerging issue in the future design of competitive learning activities.

**Conclusion**

In response to the research question posed in this study, *What are the learning preferences and motivation of different ability students for social-competition or self-competition?*, the findings indicated that: (1) Low-ability students preferred social-competition over self-competition because they emphasized on the enjoyment aspect of competitive learning activities; (2) and medium-ability and high-ability students showed similar preferences for social-competition and self-competition because they emphasized the interactive and performance aspects of competitive learning activities.

Based on the results of this study, the design of competitive learning activities could be further improved in the future. Regarding the low-ability students, the development of competitive learning might enhance the enjoyment experience, which could be realized by incorporating specific game genres. For example, students could play sports players in “management games” or heroes in “role-playing games”. In both of these students are encouraged, on the one hand, to care about their learning progress and, on the other hand, to compete against others. By doing so, the enjoyment and interactivity of competitive learning could be enhanced. Regarding the medium- and high-ability students, the development of competitive learning might enhance their awareness of learning status and strategies used to improve learning outcomes, which could be further aligned with the processes of self-regulated learning (Zimmerman & Schunk, 2001), which might contribute to their learning through planning, monitoring, reflection, and regulation, in either self-competition or social-competition.

However, due to the limitations of this study, some further efforts are required. First, this study examines students’ preferences on the choices of competitive learning but this was merely a short-term study. Students’ perceptions might be influenced by the novel effect aroused by new technologies. Thus, the long-term effects are unclear and should be addressed in the future. Second, in this study students’ ability level was measured based on examination scores in language learning, not as a measure of general cognitive ability. Thus, the application of the design framework generalizing should be limited to specifically relevant tasks, because different relationships may apply to different types of task (e.g., one requiring mathematical reasoning skills). Finally, this study emphasizes the avatar mechanism and surrogate mechanism as examples of self-competition and social-competition, which can offers a starting point to examine students’ preference, but cannot reflect all competitive models. Thus, there is a need to examine its consistency with other competitive models in future works.
Acknowledgements

The authors would like to thank the National Science Council of the Republic of China, Taiwan for their financial support of this project through contract number NSC-101-2511-S-155-004-MY2.

References


Role Blending in a Learning Environment Supports Facilitation in a Robotics Class

Ilkka Jormanainen  and Erkki Sutinen

School of Computing, University of Eastern Finland, P.O. Box 111, FI-80101 Joensuu, Finland // ilkka.jormanainen@uef.fi // erkki.sutinen@uef.fi

*Corresponding author

(Submitted June 14, 2012; Revised December 20, 2012; Accepted March 7, 2013)

ABSTRACT

The open monitoring environment (OME) uses a novel data-mining approach to enhance teachers’ pedagogical interventions in a robotics class. According to earlier studies, decision trees resulting from an open and semi-automatic data-mining process are practically useful when classifying subsequently data arising from a robotics class. The current study shows that data-mining features of the OME can be used to predict and, hence, support learning processes also with real-time data. Results show that the data-mining features of the OME are affected by the nature and amount of data when working with a small number of students. Furthermore, the results show that the robotics-class instructors are able to modify the learning environment to match to the current context in a way that goes beyond normal teacher activities in a classroom. This role blending between a teacher and a software developer in a learning environment provides a novel way to build a personalized and contextualized support environment for a robotics class.

Keywords

Robotics, Data mining, Learning environment, Teacher support

Introduction

Effective use of educational robotics requires a learning environment, which supplies open and flexible support mechanisms for the members of a learning community. Teachers working in a robotics class often face a problem in following how student groups proceed in their projects, which are usually based on problem-based or inventive learning and emphasize independent working of groups or individual students. Typically, the students work with their tasks in a cyclic process that involves planning, building, programming, and testing their robots. Groups proceed differently, perhaps being in different phases in the cycle. During ten years of running robotics activities in various technology club settings, we have seen that these characteristics of a robotics environment present difficulties for teachers planning an appropriate intervention strategy for the particular context.

Traditionally, intelligent tutoring systems (ITS) have been used to predict students’ progress in learning environments. Student modelling is usually based on theoretical assumptions about learning and a predefined sequence of actions in the learning process. However, the unpredictable nature of robotics classes establishes a need for novel approaches that allow teachers to follow up on their students’ diverse learning paths and strategies. Categorizing diverse learning strategies allows the teachers to model their students’ progress. A reasonable model is a basis to understand and, hence, to intervene on possible hindrances on students’ learning paths.

We have introduced a novel open monitoring environment (OME) to help the teacher to monitor activities and explore learning processes in a robotics class. The OME monitors students’ actions in a robotics class by automatically collecting data from the learning process, i.e., students’ interactions with educational robotics environment. The OME renders the collected data to visualizations that represent the current classroom activities. The classification rules for the collected data can be created manually based on data available in the OME database, or semi-automatically, for example, by using data mining. The OME uses an open data-mining approach (in particular, decision-tree classifiers) for modelling students’ progress (Jormanainen & Sutinen, 2012). The OME aims not only to help teachers to facilitate the students’ progress, but can also help them to account for what has been happening in the learning process. Facilitation in the context of this research means that the OME produces interpretable visualizations and explanations about the processes in the classroom that are not necessarily visible for the teacher. In this way, the OME supports the teacher in facilitating learning at the right time. This is aligned with a modern teacher’s new professional expectation as a mentor rather than as a traditional teacher.
A key element of the facilitation in the OME is to predict the possible problems that the students are facing in the learning process so that the teacher can intervene accordingly. The OME does not, however, provide information about how to intervene, but in the ideal case, it can detect the series of events that have led to the current situation. The teacher using the OME gets the support for facilitation by exploring these series of events and building his or her intervention strategies, based on the information provided by the OME.

The study presented in this paper is a part of a development research project, which aims to build tools and approaches to support teachers working in unpredictable learning environments, especially robotics classes. During the research project, we have shown that the OME helps the teachers to intervene when the students have problems in their robotics exercises (Jormanainen et al., 2012). Furthermore, we have shown (Jormanainen & Sutinen, 2012) that the open data-mining process and a proof-of-concept implementation for the data-mining module in the OME produces useful and interpretable information about the students’ progress with relatively small datasets.

In this study, the teachers used the OME in a robotics class to analyze learning process with real-time data that was generated while the students were working with the robotics environment. In fact, the learning process was understood as consecutive snapshots of equal time intervals. Each snapshot contains recorded student activities and is an opportunity for the student to progress or get stuck, thus requiring a teacher’s intervention. In our previous experiments, we analyzed the student progress data retrospectively. The main research question of this article is as follows:

*How can the OME support facilitation in a robotics class by providing the instructors learning process classifiers that are based on real-time data?*

The results from the study are two-fold. First, we identified a situation where the OME failed to produce an expressive classifier, in contradiction to our previous findings in Jormanainen and Sutinen (2012) with the retrospective analysis. In this way, the OME did not provide the needed support for the instructors’ intervention strategies. Secondly, the instructors were able to adapt the OME to meet the context-dependent requirements, and adjust the OME eventually to fulfil the monitoring needs, from the ultimate viewpoint of understanding the students’ learning difficulties. The latter result supports our findings presented previously in Jormanainen and Sutinen (2012) and Jormanainen et al. (2012), where we concluded that the OME supports a novel role-blending approach between a user and a developer.

This article is organized as follows. In the next section, we describe related work and how our approach differs from existing educational data-mining systems. Then we describe the key concepts and tools of our approach: empirical modelling (EM) and open monitoring environment (OME). After the introduction of the research design, we present the main findings from the study. Finally, we conclude the article and suggest directions for future work.

**Background**

The open monitoring environment adopts many features that can be found in existing intelligent tutoring systems, including a distributed agent architecture for collecting data and the use of data-mining techniques to predict students’ progress. The main difference between the OME and traditional ITSs is that, whereas many traditional ITSs use a theory-based approach for building the learning model, the OME starts from the empirical observations arising from the current learning situation. There are ITSs that apply an empirical approach for building the learner model, for example Wayang Outpost, a multimedia ITS for geometry by Cooper et al. (2009). However, the learning models in these systems are still at least partially predicted based on theoretical assumptions (for example, a given set of features for classifying the user’s emotional self-concept in Cooper et al. [2009]).

Data-mining tools have a recognized status as a part of ITSs and other learning environments. Usually, data mining is used to extract knowledge from e-learning systems through the analysis of the data that the users have generated (Castro et al., 2007). Most of the work in data mining in educational systems contributes to students’ assessment (e.g., Elenbogen & Seliya, 2008), learning material and course evaluation, and course adaptation based on students’ learning behaviour (Kristofic & Bielikova, 2005). Clustering and classification are the data-mining techniques most...
frequently used in learning environments (Castro et al., 2007) because of the nature of the problems that typically appear in this context.

However, two aspects limit the usability of the current data-mining approaches. First, data mining is used for inductive or deductive reasoning, i.e., inductively deriving the clusters from the available data or deductively applying the derived rules to classify data. Secondly, in most cases data mining is a black box for the learning community and its results are only implicitly visible to the users. For example, an adaptive learning environment uses learning process data to infer rules to classify learners based on their speed of progress or learning style. These rules are then used with learners or teachers, who are aware only of their existence, not of their contents. Complementary to the current mainstream uses of data mining for learning environments, we apply data mining in the OME for abductive reasoning (Ross, 2010) with transparent explanatory hypotheses. Teachers can elaborate the data-mining-generated rules to effectively facilitate the students. The rules help teachers to look backwards to the reasons behind the outcomes of the learning process, not only to efficiently support the process going forward.

The aim of an abductive problem-solving process is to find the best possible explanation for the current situation, based on data currently available. In the OME, the teacher starts the modelling process from a set of observations and generates the best possible account of the current learning process based on the observations. An abductive approach for building learning environments has been previously used by other researchers, for example by Qiu and Riesbeck (2004). However, the incremental development in Qiu and Riesbeck (2004) focuses more on contributing material during the learning process rather than on building the environment to monitor learning as the OME does.

In contrast to traditional educational data-mining applications, the mining process in the OME is transparent and results are explicitly visible to the teacher. Furthermore, the teacher is involved in the whole process, including training the data-mining algorithm, which is usually a domain expert’s or software developer’s task. Training set creation is an essential part of the process in the OME because a well-defined training set allows the data-mining algorithm to produce a well-working classifier. In the traditional data-mining application-development process, a domain expert or a software developer creates classifiers for an application with predefined data sets during the software development and deployment process. In many cases, data for training sets is processed manually, which is a very time-consuming task.

However, a teacher working in a robotics class has the best domain knowledge in his or her classroom context, and we believe that the teacher is the best expert to model the learning process in the classroom. The OME aims to provide suitable tools so that the teacher can handle the data flow originating from the robotics class and use data efficiently to produce classifiers that are aligned as well as possible with the current learning setting and pedagogical challenges.

To implement learning environments such as the OME, there is a need for tools that support abductive reasoning. Empirical modelling (EM) provides appropriate tools and methods for this. Before describing the OME environment in detail, we shall first introduce the EM tools through illustrative examples.

**Empirical modelling**

We implemented the OME by using the empirical modelling (EM) toolset (Empirical Modelling, 2012), because it is naturally well aligned with the goals of the development process. EM is a collection of principles and tools developed by Beynon, Russ, and their students at the University of Warwick, UK. EM can be used to construct computer-based models that are based on the modeller’s empirical observations about the phenomenon that is the subject of the modelling process. The modeller works with the model, the tkeden environment, by defining observables and dependencies using several different notations (Empirical Modelling, 2012).

An observable in the EM environment is an entity (such as line, string, number, window, or list of scalar values) whose current status can be inspected by the modeller. A dependency is a specific kind of relationship between two or more observables. The observables and dependencies in the EM environment are intended to serve as direct counterparts of observables and dependencies of the phenomena being modelled. The use of empirical modelling in the monitoring environment was illustrated in our previous article (Jormanainen et al, 2009). All the aspects of the current state of the model are expressed through definitions. The teacher can, in principle, manage all the details of
the model. In this way, the EM approach guides and encourages the teacher to adopt the software developer’s tasks in order to provide the right support for the students just in time.

**Open monitoring environment**

Open monitoring environment (OME) is an interactive tool that helps a teacher to facilitate students’ learning in a robotics class by providing support for the teacher about when and how to intervene in students’ work. An educational robotic setting usually produces so much data that the OME needs to provide the teacher with semi-automated tools for interacting with data. Data mining helps the teachers to synthesize the massive amount of low-level data so that it would be possible to identify more complex patterns of actions that potentially are of interest when monitoring the learning activity. It is important to note that the OME does not judge students’ learning but provides the teacher with an additional way to predict the possible moments of problems and find explanations behind them.

The OME allows the teacher to model and explore the current learning process based on data that is automatically collected from students’ interactions in the robotics environment (Figure 1). The atomic data rising from the learning process is saved to the OME database and summarised further into time slots. The teacher working with the OME synthesizes training sets for the data-mining algorithm from this data by exploring the real-time information available through the system and combines the findings with his or her own observations from the classroom. The created training sets are used to build decision-tree classifiers that are used to visualize students’ predicted workflows and situations where the particular patterns of actions may lead. Technical details of the data-mining procedure in the OME are discussed later in this article.

![Figure 1. Overall architecture of the OME](image)

We have shown previously (Jormanainen & Sutinen, 2012) that data mining, and decision trees in particular, are efficient for classifying and modelling students’ progress based on data originating from a robotics class. Based on these results, we have built a data-mining module for the OME to enable an interactive process in which the teacher is involved in building decision trees for classifying students’ progress (Jormanainen & Sutinen, 2012). In the OME, the rules are represented as decision-tree classifiers and they are explicitly opened for the teacher’s revision. The teacher can disagree with and change the automatically created rules at any time. Furthermore, the OME, in principle, allows the teacher to control and redefine all the details of the working environment. Any changes will be applied immediately in the environment, without closing it.
Data collection in the OME is based on a distributed agent architecture where software agents observe students’ actions with the robotics environment and deliver data entries with timestamps to the OME database for further processing (Jormanainen et al., 2007). Students’ learning environment consists of Lego Mindstorms RCX educational robotics sets and the IPPE graphical programming environment (Figure 2) for the Lego robots (Jormanainen et al., 2002).

Figure 2. The IPPE programming environment. Numbers 1–4 refer to Table 2

In the experimental environment, data is collected from the four major functions within the IPPE programming environment. These functions define the key steps in the robotics programming process in this context: 1. Creating individual commands to the sandbox, 2. Constructing a program from these commands, 3. Removing program lines if necessary, and 4. Uploading the program to the robot. The number of event sources was intentionally small, as our aim has been to provide a simple version of the OME for the teachers in this phase of the research project, which still involves a remarkable number of technical aspects.

The OME produces different kinds of visualizations for the current learning process. The visual representations are needed because the amount of data that a robotics environment produces can be too overwhelming for the teacher to handle as a flow of raw data (Jormanainen et al., 2009), despite the small number of attributes. Hence, the OME provides tools for enriching data to pedagogically meaningful collections. The teacher creates and manipulates rules that define how the raw data is treated and visualized. The progress (or lack thereof) can be expressed, for example, with colours or sizes of graphical elements representing the student groups.

Figure 3. Classroom view module of the OME
As the empirical modelling environment allows the user to redefine all details of a model, the teacher can, in principle, create any kinds of rules and visualizations based on them. However, complete control over an EM model requires a remarkable amount of technical skills not available in a regular classroom. In order to integrate information originating from the classroom into the teaching process, the technical environment for collecting and analyzing data needs to be simple enough, so that the lack of technical skills does not prevent the full use of the environment. At the same time, the environment needs to preserve its flexibility and provide room for the teacher to use pedagogical expertise efficiently in student progress modelling, without having predefined rules and visualizations to restrict the pedagogical process. In the experimental OME version, the students’ progress is visualized with a 2D map of the classroom (Figure 3), with the group markers showing the current situation with one of the four colours indicating the group’s progress as resolved by the current rules (Table 1).

Table 1. Classification of training data in the OME

<table>
<thead>
<tr>
<th>Class</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>Students are not progressing, but they do not seem to have any particular problems (neutral situation).</td>
</tr>
<tr>
<td>Green</td>
<td>Students are progressing without any noticeable problems.</td>
</tr>
<tr>
<td>Yellow</td>
<td>Students might be facing problems (intervention may be required soon).</td>
</tr>
<tr>
<td>Red</td>
<td>Students have experienced problems that require intervention.</td>
</tr>
</tbody>
</table>

Beside the classroom view, the OME supplies the teacher with a graphical view for the number of events in a specific time window (Figure 4). This view shows an overall progress in the class during the time window and allows the teacher to classify the events into one of the four categories according to the anticipated progress. This is a key element of data-mining process and relies on teacher’s pedagogical expertise and face-to-face experiences with the students in the classroom.

![Figure 4. Progress classifier module of the OME](image)

During the classification process, the teacher creates a training set for the data-mining module of the OME. The module uses the Weka 3 data-mining environment (Hall et al., 2009) to create a J48 decision-tree classifier for predicting the students’ progress. J48 is an open-source implementation of the popular C4.5 decision-tree algorithm (Quinlan, 1993). Table 2 presents the data sources in the programming environment and corresponding attributes in training set data. Classification of a time slot produces one line for the training set.

Table 2. Event sources in the programming environment and corresponding attributes in training set data

<table>
<thead>
<tr>
<th>Event source</th>
<th>Attribute</th>
</tr>
</thead>
<tbody>
<tr>
<td>Button 1: Add statement</td>
<td>Act_addstatement</td>
</tr>
<tr>
<td>Button 2: Add command to code</td>
<td>Act_addcommandtocode</td>
</tr>
<tr>
<td>Button 3: Remove line</td>
<td>Act_removeline</td>
</tr>
<tr>
<td>Button 4: Upload program to robot</td>
<td>Act_sendtorobot</td>
</tr>
<tr>
<td>Compiler</td>
<td>Err_comp</td>
</tr>
<tr>
<td>(no specific source)</td>
<td>Sum</td>
</tr>
</tbody>
</table>

Figure 5 shows the definitions for the training data as well as an example of the data in Weka’s arff-format. Attributes in Figure 5 are described in Table 2. In the example entry, the agents have identified one “Add command to code” event and one “Upload program to robot” event, i.e., two events within time window of one minute.
Therefore, the progress within this time window was evaluated green (students were progressing without noticeable problems).

@attribute act_addstatement numeric  
@attribute act_addcommandtocode numeric  
@attribute act_removeline numeric  
@attribute act_sendtorobot numeric  
@attribute err_comp numeric  
@attribute sum numeric  
@attribute progress_class {white, green, yellow, red}  
@data  
0,1,0,1,0,2,green

Figure 5. Training data definition and an example

Once a sufficient set of data has been classified, the teacher can create a new decision-tree classifier by pressing the “Generate tree” button in the progress classifier module (Figure 4). The OME launches a Weka data-mining application as a background process and passes the training set based on teacher’s choices to the J48 decision-tree algorithm. Without interrupting or closing the environment, Weka generates the formal description of the new classifier, based on the supplied training set. The OME transforms the description into a tree visualization (Figure 6) that is identical to the executable rules as empirical modelling definitions (Figure 7).

Figure 6. Tree visualization module of the OME

If the teacher is not confident with editing the classifier rules, he or she can always create a new classifier by continuing the classification process as students progress in their task and more data is available for a new training set. The teacher can also always browse the previous time windows and re-classify them, and try to construct a more suitable classifier in this way. This is a good approach especially at the beginning of a teaching session when training sets are too small for the J48 algorithm to build a classifier that is expressive enough. All visualizations are automatically updated through empirical modelling dependencies whenever new data is inserted to the OME database, when the teacher creates a new decision-tree classifier or when the user modifies the other EM definitions.

The IPPE, and robotics environment in general, possess a vast amount of relevant information that can be used in the future development. This information includes, for example, quantities and qualities of the students’ program code for the robots, robots’ physical construction, and even roles of the students in the groups and interaction between them. It is clear, however, that not all of these aspects can be monitored automatically. For example, observing robots’ physical construction or students’ handwritten notes is difficult as a background process in a natural learning situation where additional monitoring devices (cameras, microphones, etc.) easily interfere with the learning process.
Research design

Research methodology and research question

The experiment presented in this paper is a part of a development research project where the emphasis is on an iterative development process of the environment. We have reported the design process and initial experiments with the OME in Jormanainen et al. (2012) and in Jormanainen and Sutinen (2012). In Jormanainen and Sutinen (2012), we used the data-mining module of the OME to make a retrospective analysis of the robotics class activities. We concluded that data mining, and decision trees in particular, are effective for classifying students’ progress in the educational robotics setting, and that the open data-mining process produces useful and interpretable information about the students’ progress with relatively small datasets.

In contrast to Jormanainen and Sutinen (2012), we report in this article a study that was conducted in order to see how robotics class instructors use data-mining features to build student progress classifiers that are based on currently available, real-time empirical data. The study described in this paper focuses on the following research question:

*How can the OME support facilitation in a robotics class by providing the instructors learning process classifiers that are based on real-time data?*

Research setting

The experiment described in this paper took place in Kids’ Club of University of Eastern Finland. Kids’ Club is a combined after-school robotics and technology club and an educational technology living laboratory as a platform for designing novel educational technologies together with their intended users, particularly students and teachers. Altogether 13 club participants (between 10 and 13 years) were involved in the study. They were divided into two groups, and these groups were further divided into tree project groups of two or three students. The study was conducted individually for both groups of six or seven students. Each group spent about 30 minutes with their project.

The student groups were provided with a pre-constructed Lego Mindstorms RCX robot. They were asked to program the robot (1) to drive forward and backward for five seconds, and (2) to drive a square. The students were familiar with the basic concepts of Lego robots, as they had used Lego Mindstorms NXT educational robotics sets during the earlier club activities. However, they had not used the older Lego Mindstorms RCX or the IPPE programming environment.

Two club instructors, a 25-year-old man and a 30-year-old woman, participated to the study. Neither instructor was familiar with the OME, the IPPE programming environment, or the empirical modelling toolset before the study. The instructors had been involved in Kids’ Club activities for a school semester before the study, but they did not have a
professional pedagogical background. During the study, the instructors used the OME to observe students’ progress while instructing the students face-to-face. The instructors used the OME to iteratively create decision-tree models throughout the teaching sessions. With the first group, they started the decision-tree building process from scratch, whereas for the second group they used the final tree from the first group as a starting point. In this way, we could see how a previously built classifier would perform at the beginning of a robotics class.

Data collection

During the experiment, the instructors were using the OME to observe students’ activities in the classroom and to build decision-tree classifiers on the fly in order to see how they predict students’ progress and possible problems. The students’ interactions with the IPPE programming environment were recorded from the main functions of the programming environment (Figure 2). Altogether, 145 events were delivered over a local area network to the OME database, which was implemented as an EDDI (EDEN Database Definition Interpreter) database in the empirical modelling environment (Figure 9). Each observation contained timestamp, as well as information from which group and specific component of the IPPE programming environment the event originates from and the type of current event (user or error event).

EDDI database table structure:
observation (timestamp int, hostname int, event_type int, sender int, message int);

Examples in EDDI notation:
observation << [49462484,1,2,2,7];
observation << [49492937,2,2,1,5];
observation << [49523578,2,2,2,7];
observation << [49519937,1,2,4,11];

Figure 8. Structure and example of automatically collected event data

The summary of the events was collected and presented to the instructors in the OME environment in time windows of one minute (Figure 4). The presentation reflected a real-time progress in the classroom, and new interaction data was added to the summary whenever it became available. Based on this data, the teachers created during the session 13 different versions of decision-tree classifiers by following the workflow described above. Each decision-tree version was saved to a computer’s hard disk for further analysis. Examples of different iterations are presented in the next chapter in Figures 10 and 11.

Beside the data that the students and teachers contributed during the process, the researcher observed how the instructors used the OME, with supplementary questions during the teaching session. The questions were mainly related to how the instructors interpreted the currently available decision-tree models. Furthermore, the instructors were given minor technical guidance with the empirical modelling environment during the teaching session. After the teaching sessions, the instructors were interviewed about their opinions about the OME and how it potentially could benefit their work in robotics classes in the future.

Analysis methods and tools

The collected research material (decision-tree classifiers and supplementary material, such as field notes) was analyzed qualitatively. We analyzed the decision-tree models (n = 13) after the teaching sessions by interpreting them in the light of domain expertise that we had gathered during running robotics activities in various technology club settings for ten years. The focus of the analysis was to see how the decision-tree models have evolved during the teaching session when the amount of data in the OME database has been growing gradually while the students have been working in the robotics environment. Furthermore, instructors’ working process was analyzed in order to see how they used the OME and data-mining visualizations to support the facilitation in the learning process. The analysis was supported with extensive field notes collected during the teaching sessions.
Results

The main result of the study shows that the OME supports the teacher’s facilitation process in a robotics class by allowing the teacher to create decision-tree classifiers that can be used together with classroom visualizations (Figure 3) to interpret and explore the current progress in the classroom and to predict the possible points of problems.

The conclusion was reached by analyzing how the decision trees evolved during the teaching session while the amount of student interaction data increased gradually. The last-created decision-tree classifier (Figure 10) and the visualizations attached to it described the progress so that the instructors were able to use them to predict the problems and intervene in the students’ work before students got stuck.

Furthermore, the deeper analysis was supplemented by observations of how the instructors used the OME while helping the students and how the instructors interpreted information provided by the OME and decision-tree visualizations in particular. The main result as stated above can be further divided into two parts. At first, the OME (as prepared for this experiment) did not produce very expressive or meaningful classifiers, despite several iterations of refining training data and constructing the decision tree (Figure 9). The first version of the tree (left screenshot in Figure 9) classifies progress meaningfully, but it takes into account only a small part of the process. For example, it does not recognize actions when students are uploading the program code to robots nor possible problems occurring in this phase. This may be because the classifier was created in an early phase of the robotics class and such data has not yet been available. On the other hand, the final version of the decision tree (right screenshot in Figure 9) classifies students’ good progress based on frequencies of the “Remove line” action. In the context of the experiment (and in robotics classes in general), this hardly makes sense.

Figure 9. Two different decision-tree versions (left: the first version, right: the final version with the OME, as prepared for this experiment)

Because the instructors were not satisfied with the resulting classifiers presented above, they started to think after a while how they could reach better results with the data available. When reclassification of the existing data did not produce the desired result, they started to modify the rules produced by the data-mining algorithm in the rule window of the OME (Figure 7). After carefully examining the decision tree available at that time, the instructors disagreed with some choices made by the J48 algorithm. For example, they changed variable `sum` to variable `add_statement` in one of the conditions (tree nodes) in order to achieve better results. When even these changes did not improve the result enough, the instructors started to explore more deeply how the data-mining process works in the OME.

The researcher had explained earlier the basic principles of the classification and training set creation in the OME. By using this knowledge, the instructors were able to detect a possible bottleneck in the process. When the OME creates a training set, it divides sums of individual events with the number of the groups within a time window. For example, if three student groups had produced 12 events in total within a time window, the training set was supplied with an average of four events. The instructors concluded that this might prevent a proper classification due to the size or other characteristics of data sets available in the current robotics class, that is, the sums of data items were so small that taking averages from these sums led to training sets with inappropriate values for the J48 algorithm.

Guided by the researcher, the instructors explored the functionality and definitions of the OME through the empirical modelling environment. As part of the modelling process, the EM environment allows a user to query and modify the state of the current observables, functions, and procedures with an interactive `tkeden` interpreter.
After exploring the OME definitions for a while, the instructors identified a procedure that is used to create the training set based on the user’s choices. Furthermore, the instructors identified a block in the procedure (three lines of the EM definitions) that calculates the averages. The instructors opened an appropriate code file in a text editor and removed the unwanted block. After that, they applied the modified procedure to the OME without interrupting the environment. It is notable that the instructors were not previously familiar with the EM environment or the syntax of the EM definitions. Hence, the researcher gave technical advice to them to complete the task. As a result, the data-mining module in the OME started to produce significantly better results (for clarity, the tree in Figure 10 has been redrawn based on the original OME visualization).

Figure 10. The decision tree after modifications of the OME

At the first sight, the classifier in Figure 10 is more expressive and it describes the learning process better than the final result with the original OME environment (Figure 9). For example, progress is classified as the green category (students are progressing without any noticeable problems) in N6 when there are no errors and the students have removed only few lines from the code they are working with. This indicates that they have created commands, constructed robot code from these commands, and finally uploaded the program to the robot successfully (error events occur when there are problems with compiling or uploading the code).

A thorough interpretation of the decision tree reveals that the classifier exposes patterns of actions that would be otherwise hard to predict. For example, the classifier indicates possible problems arising (yellow) in the node N8 when there are no errors (N4) and the students have removed more than two lines from the code (N5), but they have applied “Add command to code” functionality just once or not at all (N7). This is not a part of a typical workflow with the IPPE programming environment, but the situation is, indeed, possible to reach as the IPPE allows editing code straight to the code editor. In the current research setting, the students had, however, reached the situation by removing the initial program code lines for the body of the robot’s program that appeared in the editor during the start-up. This is clearly an undesired action, because it makes compiling the program code impossible and teacher’s intervention needed.

After the teaching session, the instructors analyzed the final classifier based on their experiences from the classroom. They were able to trace the tree from the root node and justify the choices made by the J48 algorithm. Furthermore, they agreed that the tree gives a deeper view for the individual group’s progress in this particular robotics environment than the trees that the original OME version produced (Figure 9). Finally, the instructors concluded that the tree model could serve as a suitable starting point when preparing the OME for future robotics classes.
Discussion

The findings from this study contradict, to some extent, our previous results (Jormanainen & Sutinen, 2012), when we tested the data-mining capabilities of the OME with small data sets. It is notable that in the current experiment, an average size of a data set in a time window was 20% smaller than that presented in Jormanainen and Sutinen (2012). This may have had a negative effect for the functionality the J48 decision-tree algorithm in the first place. On the other hand, this strengthens our assumption that data-mining process in the OME is very sensitive to the nature of data collected from the learning process, as well as to user’s personal preferences and pedagogical expertise. This makes the classifiers highly context-dependent, and there is a need to provide the teacher with the tools that allow him or her to deeply control the details of the student modelling process.

The interpretability and pedagogically meaningful constructions of decision-tree classifiers are even more important aspects in the OME than the traditional measurements, such as the accuracy and precision ratio that are usually used to evaluate the data-mining applications. It is evident that small data sets in the experiment interfered at first with the data-mining process so that the resulting classifiers were not as expressive or interpretable as those presented in our previous study (Jormanainen & Sutinen, 2012). However, more interesting and unexpected results rose from the instructors’ working process with the OME. The most interesting finding was that the instructors were independently able to identify the problem in the functionality of the OME and explore the construction of it with the supplied tools. More importantly, they were able to apply the changes to the core OME definitions so that the environment started to produce better classification results in the current learning setting. This process goes beyond a teacher’s traditional role and it is well aligned with the principles of the role-blending approach as described in Jormanainen et al. (2012).

If the modifications that the instructors made during the experiment are considered strictly in the light of correctness of the program, it is notable that the modifications made the OME to work wrong. As a result, the training set based on cumulative sums of events is used to create a classifier to predict the progress of individual groups, and we can say that in a general software development project this is not appropriate. However, in the context of this particular robotics class, the “wrong” behaviour of the OME was desired at least when only small data sets that were available during the experiment. This shows that a learning environment can be constructed by using unconventional solutions if it helps to support the unpredictable nature of the learning setting, as long as the changes are manageable by the end-user (in this case, the teacher).

Conclusion

We have presented a novel monitoring environment that facilitates teachers’ intervention in educational robotics classes. The environment uses data mining to classify events rising from the classroom and to predict students’ progress. Unlike traditional educational data-mining applications, the open monitoring environment allows the teacher to create training sets for the decision-tree algorithm based on real-time data. The created classifier is visualized for the teacher and forthcoming events are classified based on this decision tree. If not satisfied, the teacher can iterate the process with more data or manipulate the current classification rules to match them as well as possible to the current learning scenario.

We demonstrated previously that decision trees are efficient when classifying data originating from the robotics classroom (Jormanainen & Sutinen, 2012). In this paper, we have presented a study, which shows that a) the selected J48 decision-tree algorithm is sensitive for the nature and amount of data when working in small robotics settings, and b) instructors in a robotics class are able to modify on-the-fly even the core definitions of the monitoring environment to contextualize the environment to suit to the current conditions (i.e., a small number of groups and not enough data). This process with role blending between the teacher and the software developer differs radically from the traditional approaches with static roles and predefined working environments. Effects of group size and amount of data must be studied further in a full-scale robotics class (20–30 students working in parallel). Another aspect for the further studies is to allow the teacher to use arbitrary time windows in the classification instead of using a fixed time window of one minute. Letting the teacher decide suitable lengths for the time windows could enable deeper interpretation of the learning process when the specific events in the classroom can be connected to each other over the longer periods of time.
Acknowledgements

We would like to thank Dr. Meurig Beynon (University of Warwick, UK) for his comments for the manuscript.

References


Towards an Enhanced Learning Management System for Blended Learning in Higher Education Incorporating Distinct Learners’ Profiles

Sofia Balula Dias* and José Alves Diniz

University of Lisbon, Faculty of Human Kinetics, Estrada da Costa, 1495-688 Cruz Quebrada, Lisbon, Portugal // sbalula@fmh.utl.pt // jadiniz@fmh.utl.pt

*Corresponding author

(Submitted December 18, 2012; Revised March 17, 2013; Accepted May 3, 2013)

ABSTRACT

In a blended education context, Learning Management Systems (LMS) can be thought to integrate collaborative and interactive learning activities; this, however, requires a strong institutional and sociocultural commitment from all stakeholders. Consequently, an empirical study that aims at identifying learners’ profiles and uses them as an optimization feedback-like process to the LMS towards effective blended (b-)learning was adopted here. The latter involved 36 undergraduate students with variant b-learning activity at a public Higher Education Institution. In a synergetic combination of qualitative and quantitative evidences, semi-structured face-to-face interviews were conducted and validated, and a systematic multivariate content analysis was articulated. Results revealed three distinct students’ profiles oriented to interactive learning environment, Information and Communication Technologies (ICTs) teachers’ beliefs, and students’ training. Under this multifaceted scenario, a rethinking of the LMS within the b-learning environment could be approached through the enhancement of interactivity, fostering users’ ICT acquaintance, and incorporating further training.

Keywords

Learning Management System, Blended learning, ICT knowledge, Higher Education, Learners’ profiles

Introduction

World Wide Web phenomena, in general, and the Web-based Instruction, in particular, can open up (re)new opportunities for the development of micro-educational environments. The current paradigm shift from traditional educational environments to online educational environments in higher education can also be seen as a challenge to create an active and interactive learning environment, one which gives the learner opportunity to engage and think in multiple ways (Bonk & Reynolds, 1997). In fact, there is a variety of electronic learning (e-learning) environments that consider and combine different Information and Communication Technologies (ICTs) tools and instructional strategies; nevertheless, blended learning (b-learning), which combines face-to-face and online learning, has been considered as “the most common mode of e-learning” (Bates & Sangrà, 2011a, p. 42). Actually, b-learning meets multiple and differentiated instructional online activities, therefore it has the potential, human and technological, of accommodating students with distinct learning needs. In this sense, ICT can potentially be used: i) to support creative and collaborative learning activities in distance education, which can easily emerge from a “meeting of minds”, where clusters of students can both work and discuss ideas online (Wheeler, 2005, p. 151); and ii) to create a learning culture that supports the students’ individual processes. Multi-stakeholders (e.g., alumni, faculty and administrators), nevertheless, should be actively engaged in this process to encourage a more holistic, comprehensive, co-constructive and globally convergent attitude.

Accordingly, some studies suggest that the development of a harmonious and effective online course depends on feedback interventions and motivational strategies (Fisher & Baird, 2005; Simonson, 2005). In addition, pedagogical design, assessment activities, as well as feedback, seem to represent key features to validate the online formative assessment in higher education (Beatty & Gerace, 2009; Gikandi et al., 2011). In turn, Simonson et al. (2011) have argued that rethinking and reorganizing online teaching and learning dynamic through emergent phenomena, such as group interaction, collaboration and teamwork, requires the establishment of complex/golden roles in the process of higher-order collaborative learning and construction of knowledge. In fact, an effective b-learning environment enables users to participate in the co-creation, produsage, of knowledge through social and technological affordances, and to promote skills and competencies, such as creativity, adaptability, communication and higher-order thinking (Redecker et al., 2009).
To improve the quality/optimization of a b-learning process it seems reasonable to take into account both educational resilience and interpersonal engagement, in order to understand constructivist and sociocultural core principles. Furthermore, technological systems, and their analogous tools, can be valuable educational instruments in b-learning, depending on the way they are used. From the latter perspective, Learning Management Systems (LMSs) have been structured, as a means to manage teaching and learning activities in an online learning environment, such as that of b-learning (Black et al., 2007). A typical LMS embeds agents that belong to an interactive learning environment assisted by mediating tools that support, for example, inter/intra-action, collaboration, training, communication and sharing information amongst the LMS users. Regrettably, some LMSs are primarily used as a tool set for information distribution and administrative effectiveness rather than a system with potential to improve teaching and learning activities (Black et al., 2007; Kvavik et al., 2004). To achieve the latter, the role of the LMS users should be taken into consideration. In particular, in a process that focuses on learning, it is particularly important to provide an enriched educational experience to all students in an environment that is supportive and inclusive. Thus, in order to prevent these specific technological and emotional weaknesses, it seems essential to evaluate the multidimensional and dynamic usability nature of a LMS, i.e., effectiveness, learnability, flexibility and the users’ attitude (Shackel, 2009). From this perspective, some questions arise, regarding the effect of the LMS structure on the students’ learning and vice versa, i.e., the perception of the role and functionality of a LMS in their effective learning and social engagement. In this line, for example, Jones et al. (2008) from an academic perspective, specifically in coursework, suggest that the use of internet-based LMS is not necessarily correlated with student’s satisfaction. Obviously, there is a need to further examine the way LMS is affected by the learners’ profiles, in an effort to incorporate the latter information in the enhancement of its design. This sets the basic rationale for the current study, as explained in details in the following section. Subsequently, the methodology used, along with study design, participants’ characteristics, data collection and implementation issues are included in the succeeding section. Next, results from the research study are reported, followed by discussion about the optimization processes that need to be considered for the further enhancement of LMSs according to the learners’ profiles previously presented, giving also some future research directions. Finally, the paper ends with the main research conclusions.

Research rationale

The rational of this study was to assess students’ needs and to identify and understand their profiles, in order to optimize the quality of online teaching and learning, in a specific b-learning context within a LMS environment. In this way, the learners’ profiles could shed light upon some specific issues concerning the key factors of the LMS structure and reveal the weaknesses and drawbacks that need further consideration.

Usually, a LMS under the b-learning perspective includes structural parts that assist a managerial administration of the knowledge carriers, realized through an interactive learning environment (see Figure 1). Apparently, a set of mediators is also necessary to assist the handling of the interactions of the users with the LMS. To this end, mediating tools that, amongst other, foster users’ inter/intra-actions, collaboration, training, communication, data logging and sharing information are included in the LMS (see Figure 1). The role of the LMS is to increase the efficiency of the b-learning that is received by its users, i.e., learners (explicitly) and teachers (implicitly). If we consider the learners’ profiles as an additional source of information, then optimization processes could take place that could be fed into the LMS and, by affecting some structural elements, could increase the efficiency of the outputted b-learning to the receivers. This is schematically presented in Figure 1, where the depicted block-diagram shows an ordinary LMS with its inputs and output, combined with an additional branch (included in the dashed rectangular) that controls the LMS output (efficiency of b-learning).

The motivation to focus upon the learners’ profiles as a first priority was drawn from the fact that the whole process of b-learning is actually designed for them and they are the main receivers of knowledge through their interaction with the LMS environment (individually and/or within groups). Hence, their role in the latter could be approached twofold, i.e., as simple users (passive mode) and as contributors to the enhancement of the LMS functionality by identifying points of improvement and expressing their opinions after its use (active mode). Considering the information from the active mode, the LMS becomes more functional and encourages learners’ participation and contributions. In a step further, teachers’ profiles could also be incorporated, yet with a reference to the learners’ ones, as there is a kind of cause-effect dependency among them (see further analysis on this issue at the discussion section).
As the learners in the active mode act as a kind of feedback to the LMS, their profiles were approached with care, eliminating any potential subjective bias, fostering the objective dimensions that lie in their responses. To this end, a specific methodological approach was adopted that handles multivariate categorical data acquired from the learners, as described in the succeeding section.

Methodology

The methodology adopted here for acquiring learners’ profiles combines both qualitative and quantitative evidence in a synergistic way, i.e., using qualitative data for understanding the rationale and theory underlying relationships revealed in the quantitative data and, in turn, avoiding any false impressions and subjective interpretations from qualitative data.

The qualitative evidence was based on a semi-structured face-to-face interview of the learners in the active mode, with questions previously validated from experts in the field. The interviews were structured in four distinct parts, hereafter designated as Categories. Category 1 enabled the characterization of the communication tools used in the LMS environment (*LMS Moodle tools*), Category 2 aimed to analyse the potential benefits of the LMS concerning the collaborative and interactive network (*Potential strengths*), Category 3 aimed to understand the concerns about the use of LMS (*Weaknesses*), and Category 4 intended to retrieve data regarding the students’ expectations to future LMS usage (*Students’ suggestions*). The combination of the qualitative data from Categories constructs the data space, which reflects information about the students’ satisfaction with the use of the b-learning in the LMS environment, their perception of it and the instructional strategies/tools were used. The qualitative data space is then subjected to quantitative analysis using the content analysis software MAX Qualitative Data Analysis (MAXQDA, http://www.maxqda.com) to develop a classification/coding system, i.e., construction of the quantitative data space. The latter, then, was coded into a large number of subcategories, reflecting both the research questions and the themes which emerged from a close analysis. The coded data produced were statistically explored and articulated using the statistics analysis software SPSS18 (http://www-01.ibm.com/software/analytics/spss/). A Multiple Correspondence Analysis (MCA) was then conducted, since it is considered a useful technique for the structural analysis of multivariate categorical data and is also suitable for reducing the dimensionality of the original...
categorical variables set (Benzecri, 1992). In fact, the MCA is a factorial method that displays categorical variables in a property space, which maps their associations in two or more dimensions. From a table of \( n \) observations and \( p \) categorical variables, describing a \( p \)-dimensional cloud of individuals \( (p < n) \), the MCA provides orthogonal axes to describe most of the variance of the whole data cloud. The fundamental idea is to reduce the dimensionality of the original data thanks to a reduced number of variables (factors), which are a combination of the original ones. The MCA is generally used as an exploratory approach to unearth empirical regularities of a dataset (Benzecri, 1992).

The characteristics of the participants in the study, along with some implementation issues are described in the succeeding subsection.

### Participants’ characteristics & Implementation issues

In order to get a holistic overview of how the LMS has been used to potentiate the b-learning, the perspectives of the students of five different courses (Sport Sciences, Ergonomics, Dance, Sport Management and Psychomotor Rehabilitation) offered by a public Higher Education Institution (HEI), i.e., Faculty of Human Kinetics, University of Lisbon (Portugal), were acquired via the aforementioned interviews. These courses reflect the core of the curriculum of the specific HEI, involve a high number of students each academic year and offer a distinct LMS-based b-learning environment.

In an effort to select the learners’ sample as objectively as possible, from a total of around 800 LMS undergraduate users, two distinct groups were formed according to the frequency of LMS usage, i.e., the most active students (Group 1, 18 students) and the less active ones (Group 2, 18 students), resulting in a total number of \( n = 36 \) selected undergraduate participants. In this way, the diversity between the learners’ groups is taken into account in the data space (deviant sample) [27]. This selection was based on the available data from LMS website, considering three usage indicators that refer to: (i) number of views, (ii) number of contributions and (iii) total activity in the LMS environment. In general, the more and less active students were selected in each course, according to number of hits, i.e., total access, allowing to differentiate between Group 1 (Median; Inter-quartile Range [75%-25%] = 2856.5; 1290.3) and Group 2 (Median; Inter-quartile Range [75%-25%] = 737; 623). It should be noted that the distribution of the number of participants for each LMS-based course was similar across the two groups. Moreover, the concept of a random sample was not adopted here, since although it provides the best opportunity to generalize the results to the population, is not the most effective way of developing an understanding of complex issues relating to human behavior (Marshall, 1996).

From the 36 undergraduate participants involved in the study, 61% of them was female, while their age ranged from 18 to 48 yrs (mean±standard deviation = 22.05±5.44 yrs). Data were collected at the end of the first semester of the academic year 2010/2011 (January), whereas the 53% of these participants (19) started to use the LMS (i.e., made their first access) in the academic year 2009/2010. In fact, 33% (12) of them made their first entry in September 2009 and 47% (17) reached their peak activity during the 5-month period, from January to May 2011 (see Table 1). All participants used b-learning via LMS Moodle (http://moodle.org/) for at least 6 months.

<table>
<thead>
<tr>
<th>Categories/Time period</th>
<th>Number of students (n = 36)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
</tr>
<tr>
<td><strong>First access in LMS (academic year)</strong></td>
<td></td>
</tr>
<tr>
<td>2008 / 2009</td>
<td>2</td>
</tr>
<tr>
<td>2009 / 2010</td>
<td>19</td>
</tr>
<tr>
<td>2010 / 2011</td>
<td>15</td>
</tr>
<tr>
<td><strong>First entry in a subject (month, year)</strong></td>
<td></td>
</tr>
<tr>
<td>September, 2008</td>
<td>2</td>
</tr>
<tr>
<td>September, 2009</td>
<td>12</td>
</tr>
<tr>
<td>October, 2009</td>
<td>5</td>
</tr>
<tr>
<td>November, 2009</td>
<td>1</td>
</tr>
<tr>
<td>January, 2010</td>
<td>1</td>
</tr>
<tr>
<td>September, 2010</td>
<td>10</td>
</tr>
</tbody>
</table>
The acquired interviews were initially audio-recorded and followed by a verbatim transcription, i.e., thirty-six protocols were obtained through the interview transcriptions (Student01-Student36) and formatted so they could be fed in the MAXQDA software. Three interviews were also chosen randomly for the purpose of testing the coding reliability. Furthermore, the first three dimensions of MCA that explain most inertia and have high eigenvalues were kept. This follows what Benzécri (1992) suggests, noting that this limit should be fixed by user’s capacity to give a meaningful interpretation to the axes he keeps by checking eigenvalues and the general meaning of dimensions. In addition, the threshold value for the weighted correlations adopted in the MCA to perform dimensionality reduction was equal to 0.5, so only the important variables to be considered per dimension. Finally, the combination of qualitative and quantitative analyses established a methodological triangulation (Denzin, 1970; Fielding, 2009). In the section that follows, analysis results are described in details.

**Results**

The content analysis using MAXQDA software performed in the interviews allowed to associate a set of items, i.e., Subcategories, to each Category (defined in Methodology Section). Subcategories in each Category emerged as the most important topics from the interviews corresponding to Text Units (TU). The latter were codified as a unit of meaning, a word, a phrase or a paragraph, using the semantic criteria in a hermeneutic interpretation. The emerged Subcategories per Category along with the corresponding total TU are tabulated in Table 2.

<table>
<thead>
<tr>
<th>Category</th>
<th>Subcategory</th>
<th>TU Total</th>
<th>TU (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LMS Moodle tools</strong></td>
<td>Quiz</td>
<td>6</td>
<td>2.38</td>
</tr>
<tr>
<td></td>
<td>Wiki</td>
<td>10</td>
<td>3.97</td>
</tr>
<tr>
<td></td>
<td>Webmail</td>
<td>24</td>
<td>9.52</td>
</tr>
<tr>
<td></td>
<td>Forum</td>
<td>31</td>
<td>12.30</td>
</tr>
<tr>
<td></td>
<td>Label</td>
<td>13</td>
<td>5.16</td>
</tr>
<tr>
<td></td>
<td>Files/resources</td>
<td>50</td>
<td>19.84</td>
</tr>
<tr>
<td></td>
<td>Chat</td>
<td>13</td>
<td>5.16</td>
</tr>
<tr>
<td></td>
<td>Glossary</td>
<td>8</td>
<td>3.17</td>
</tr>
<tr>
<td></td>
<td>Assignment</td>
<td>44</td>
<td>17.46</td>
</tr>
<tr>
<td></td>
<td>Linkability to other systems</td>
<td>53</td>
<td>21.03</td>
</tr>
<tr>
<td></td>
<td><strong>total</strong></td>
<td><strong>252</strong></td>
<td><strong>100.00</strong></td>
</tr>
<tr>
<td><strong>Potential strengths</strong></td>
<td>Teacher-student interaction</td>
<td>36</td>
<td>14.52</td>
</tr>
<tr>
<td></td>
<td>Courses at postgraduate level</td>
<td>2</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>Sharing information</td>
<td>17</td>
<td>6.85</td>
</tr>
<tr>
<td></td>
<td>Content repository</td>
<td>59</td>
<td>23.79</td>
</tr>
<tr>
<td></td>
<td>Organization</td>
<td>20</td>
<td>8.06</td>
</tr>
<tr>
<td></td>
<td>Efficiency in learning</td>
<td>9</td>
<td>3.63</td>
</tr>
<tr>
<td></td>
<td>Communication</td>
<td>31</td>
<td>12.50</td>
</tr>
<tr>
<td></td>
<td>Accessibility</td>
<td>19</td>
<td>7.66</td>
</tr>
<tr>
<td></td>
<td>Usability</td>
<td>28</td>
<td>11.29</td>
</tr>
<tr>
<td></td>
<td>Self-regulated learning</td>
<td>15</td>
<td>6.05</td>
</tr>
<tr>
<td></td>
<td>Collaboration</td>
<td>10</td>
<td>4.03</td>
</tr>
<tr>
<td></td>
<td><strong>total</strong></td>
<td><strong>248</strong></td>
<td><strong>100.00</strong></td>
</tr>
</tbody>
</table>

| Weaknesses                | Students’ low ICT knowledge      | 63       | 22.99   |

| October, 2010             | 5                                | 13.89    |
### Technical issues

<table>
<thead>
<tr>
<th>Category</th>
<th>TU</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online safety</td>
<td>8</td>
<td>2.92</td>
</tr>
<tr>
<td>Age and sociocultural differences</td>
<td>20</td>
<td>7.30</td>
</tr>
<tr>
<td>Teachers’ low ICT knowledge</td>
<td>46</td>
<td>16.79</td>
</tr>
<tr>
<td>Lack of time</td>
<td>8</td>
<td>2.92</td>
</tr>
<tr>
<td>High number of students</td>
<td>17</td>
<td>6.20</td>
</tr>
<tr>
<td>Teachers’ beliefs, subject matter</td>
<td>43</td>
<td>15.69</td>
</tr>
<tr>
<td>Techno-pedagogical knowledge</td>
<td>33</td>
<td>12.04</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>274</td>
<td>100.00</td>
</tr>
</tbody>
</table>

### Online safety

<table>
<thead>
<tr>
<th>Category</th>
<th>TU</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-tool systems</td>
<td>20</td>
<td>13.07</td>
</tr>
<tr>
<td>Content reorganization module</td>
<td>20</td>
<td>13.07</td>
</tr>
<tr>
<td>Formative feedback</td>
<td>19</td>
<td>12.42</td>
</tr>
<tr>
<td>Faculty training</td>
<td>34</td>
<td>22.22</td>
</tr>
<tr>
<td>Students’ ICT training</td>
<td>13</td>
<td>8.50</td>
</tr>
<tr>
<td>Assessment tasks</td>
<td>34</td>
<td>22.22</td>
</tr>
<tr>
<td>Others</td>
<td>13</td>
<td>8.50</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>153</td>
<td>100.00</td>
</tr>
</tbody>
</table>

### Note.
Results for each individual Category are described next.

#### Category 1: LMS Moodle tools

As specified in Table 2, in the LMS Moodle tools category, 252 TU in total were codified. The Subcategory “Linkability to other systems” was the dimension most valued by the students (TU = 53, 21%), followed by the Subcategory “Files/resources” (TU = 50, 20%) (Table 2). In this context, the use of other systems (e.g., class mail, blogs) is reinforced by several of the interviewed students, for instance: “Class mail is restricted to students and teachers do not have access to it. We usually use the LMS to upload documents, sometimes to communicate with teachers. The LMS is an official system and class mail is a more informal online environment, which allows sharing some information or class notes privately” (Student05). Therefore, the LMS seems to be more used as a content repository rather than a collaborative learning environment. Student36 reinforces this perspective when he states: “Basically, I use the LMS to download subject contents, pdf documents, lecture notes or slides.”

#### Category 2: Potential strengths

In turn, concerning the Potential strengths category, (248 TU in total, Table 2), the Subcategories most valued were “Content repository” (TU = 59, 24%) and “Teacher-student interaction” (TU = 36, 15%), as indicated in Table 2. Similarly, students exposed relevant advantages in using a distance learning system, e.g., to support asynchronous collaborative activities and/or to provide meaningful opportunities of self-regulated learning (Table 2). For instance, some students revealed: “The LMS is very important to download documents, slides, study notes, which help us a lot […] it is extremely convenient and easy to use.” (Student21), and “I usually use the LMS to download documents, music, videos […] information that we really need to write our essays.” (Student07). It was also possible to infer that students consider the LMS Moodle as a social media tool; as reported by one student: “[…] the LMS is an easy way to communicate with teachers, e.g., to submit and share documents.” (Student04).

#### Category 3: Weaknesses

In the Weaknesses category, 274 TU in total were classified (Table 2). The Subcategory “Students’ low ICT knowledge” includes the highest TU values (TU = 63, 23%), followed by the Subcategory “Teachers’ low ICT knowledge” (TU = 46, 17%) (Table 2). Accordingly, the lack of ICT knowledge of both students and teachers tends to appear as an important limitation. In this context, two interviewees considered that: “In terms of interaction, I do
not know how to use correctly some tools available online learning platform. I know that the LMS has many potential tools, like chat rooms or discussion forum, which allows us to interact and share information with colleagues. But, I only use the LMS to download slides presentations, lecture notes or supplementary texts” (Student11); “Some teachers indicate potential advantages in the use of the LMS and others are much more reluctant to use it, because they only post presentation slides (used in face-to-face sessions) in the platform. I think that the main reason why they prefer to use traditional teaching activities is technophobia, the lack of time, and some do not know how to use the online tools.” (Student23).

Category 4: Students’ suggestions

Regarding the Students’ suggestions category, referring to optimization of the LMS use, 153 TU in total were classified (Table 2). The Subcategory “Faculty training” and the Subcategory “Assessment tasks” symbolize the highest amount of TU coded in this category (TU = 34, 22%) (Table 2). These issues were highlighted by the perspective of a student when he states: “I think that they [the teachers] need to develop technological knowledge, probably through faculty training, and so that they can develop activities using different tools, such as wikis, discussion forums [...] or create learning sequences, such as a Learning Activity Management System activity.” (Student32). Still another student indicated that: “The use of the LMS should be mandatory, teachers should stimulate the use of the platform by defining assignments, the participation in discussion forums (as mandatory) consequently, students would be more motivated to use the LMS.” (Student34).

MCA Dimensions identification

In Table 3, the results of the MCA are presented, using all categorical variables considered in the content analysis of interviews, in order to reduce the dimensionality of the Subcategories reported in Table 2 and better understand their relations. Thus, the MCA technique allowed for the clustering of the Subcategories into three different students’ profiles regarding the LMS use, namely *Interactive learning environment* (Dimension 1), *ICT teachers’ beliefs & differentiation* (Dimension 2), and *Students’ Training* (Dimension 3), considering those Subcategories per Dimension that their weighted correlation exhibits values > 0.5 (denoted in bold in Table 3). The Cronbach’s alpha (\( \alpha \)) was considered to determine the reliability and to assess the internal consistency of the dimensions presented above (Benzecri, 1992), and the obtained values were .96, .94 and .93, for Dimensions 1, 2 and 3, respectively. This indicates a good internal consistency and reliability in the definition of the learners’ profiles. Based on these findings, the latter and the corresponding optimization processes (Figure 1) are discussed in detail in the succeeding section.

<table>
<thead>
<tr>
<th>Variable (Subcategory)</th>
<th>Dimension 1</th>
<th>Dimension 2</th>
<th>Dimension 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Webmail</td>
<td>1.657</td>
<td>.112</td>
<td>.056</td>
</tr>
<tr>
<td>Chat</td>
<td>.876</td>
<td>.164</td>
<td>.387</td>
</tr>
<tr>
<td>Teacher-student interaction</td>
<td>1.363</td>
<td>.323</td>
<td>.177</td>
</tr>
<tr>
<td>Sharing information</td>
<td>.861</td>
<td>.433</td>
<td>.028</td>
</tr>
<tr>
<td>Self-regulated learning</td>
<td>.650</td>
<td>.014</td>
<td>.342</td>
</tr>
<tr>
<td>Accessibility</td>
<td>1.192</td>
<td>.239</td>
<td>.341</td>
</tr>
<tr>
<td>Efficiency in learning</td>
<td>1.225</td>
<td>.056</td>
<td>.001</td>
</tr>
<tr>
<td>Teachers’ beliefs, subject matter</td>
<td>.060</td>
<td>5.637</td>
<td>.106</td>
</tr>
<tr>
<td>Lack of time</td>
<td>.036</td>
<td>.356</td>
<td>.619</td>
</tr>
<tr>
<td>Linkability to other systems</td>
<td>.228</td>
<td>.436</td>
<td>2.075</td>
</tr>
<tr>
<td>Glossary</td>
<td>.192</td>
<td>.030</td>
<td>.647</td>
</tr>
<tr>
<td>Students’ ICT training</td>
<td>.291</td>
<td>.097</td>
<td>.894</td>
</tr>
<tr>
<td>Usability</td>
<td>.334</td>
<td>.496</td>
<td>1.643</td>
</tr>
<tr>
<td>Collaboration</td>
<td>.030</td>
<td>.091</td>
<td>.376</td>
</tr>
<tr>
<td>High number of students</td>
<td>.021</td>
<td>.138</td>
<td>.122</td>
</tr>
<tr>
<td>Label</td>
<td>.323</td>
<td>.188</td>
<td>.174</td>
</tr>
</tbody>
</table>
Note. The Subcategories per Dimension that their weighted correlation exhibits values > 0.5 are denoted in bold. The values exceeding 1.0 have no physical meaning, since they are produced from some missing values in the qualitative data. To this end, they should only be considered as indicators of strong dependence of the specific variable to the corresponding dimension.

Discussion

Learners’ profiles & optimization processes

Interactive learning environment (Dimension 1)

This dimension explains the type of learning environment valued by the students. Statistical results (eigenvalue = 10.510, inertia = .618) suggest that there is a highly positive relationship between the use of distinct – synchronous and asynchronous – communication tools (Webmail, Chat), the benefits of interaction (Teacher-student interaction, Sharing information), the sustainable education (Self-regulated learning), and the user-friendliness (Accessibility, Efficiency in learning).

In fact, interactive environments are determinant in distance learning as they may condition the success of the learning outcomes and the quality of online learning per se (Abrami et al., 2011; Muirhead & Juwah, 2004). Some researchers demonstrated that the creation of a learner-centered LMS implies particular interaction relations associated with online learning, namely, learning-interface, learner-self, learner-content, and learner-learner (Chou et al., 2010; Hirumi, 2009). Thus, the features of LMS will allow an adjustable and dynamic ecosystem that can integrate different interactive learning activities. Based on the students’ responses, the empowerment and continuous improvement of LMS interactivity may result in higher levels of students’ satisfaction; in their own words: “I believe that some teachers are more comfortable using interactive tools, such as wikis, assignments, forums or a chat than others […] depends on the subjects, but we have more motivation and high-interest for interactive activities; I think that the learning process is, this way, easier and more attractive.” (Student17). In this sense, interactive environments with a diversified and integrated approach may enrich and reinforce students’ intrinsic interest in academic online activities, namely to a motivational process promoting self-regulated learning. As a matter of fact, various authors have also demonstrated that the structure of learning communities is a particularly important issue that should be considered to develop processes of higher-order thinking and contextual learning (e.g., through social interactions) (Redecker et al., 2009; Zhao & Kuh, 2004). For instance, more recent interactive technologies, such as mobile social computing, can be integrated to encourage student communication, collaboration and creativity (e.g., micro-blogging, life-streaming, social tagging, podcasting, social networking, media sharing) (Gupta et al., 2009; Redecker et al., 2009). This can also contribute to understand and respond to both students’ social needs and cultural diversity. According to McGuire (1996), interactive learning environments represent a blend of both multimedia and hypertext, which integrate, for example, analogous/associative characteristics, accessibility, linkability, intuitiveness, and nonlinear organization. As a result, the combination effect of nonlinear, multisensory, and multimodal interactive systems seems to offer strong potential to expand b-learning scenarios.

ICT teachers’ beliefs & differentiation (Dimension 2)

This dimension recognizes the importance of teachers’ ICT knowledge in the LMS usage being solely constituted by one subcategory (Teachers’ beliefs, subject matter, eigenvalue = 8.846, inertia = .520).

Teachers’ beliefs, particularly sociocultural beliefs, seem to represent a large role on how distance learners from different parts of the world interact with teaching and learning systems. In fact, certain internal constraints (e.g., teachers’ beliefs, teachers’ self-efficacy, teachers’ attitudes) and external constraints (e.g., access, training, local support) were identified as relevant barriers that influence the teachers’ ICT implementation efforts (Ertmer, 1999). Accordingly, both cultural identities and thinking processes have been highlighted as important obstacles to the integration of ICT in the educational context (Richards, 2004; Watson, 2001). Equally important, some studies have pointed out that only a few teachers are prepared to integrate ICT into their teaching activities, even though there is an increasing awareness of teachers as to the value of training as to ICT use (Dawes, 1999; Kirkup & Kirkwood, 2005). In students’ point of view, some differences in teachers’ behaviour are associated with ICT knowledge and
intrinsic motivation. An interviewee stated that: “I think that some teachers are more familiar with the technology, and others just do not use the tools and resources that are available in the LMS [...] they need to be more self-confident about using the LMS for teaching-learning activities” (Student32). There is also evidence that, the disciplinary differences are critical factors in the design and improvement of online courses (Arbaugh et al., 2010; Smith et al., 2008). For instance, distance learning in applied disciplines (e.g., Engineering, Nursing, Education) appears to be more diversified and more geared towards a community of practice, compared to the pure disciplines (e.g., Natural Sciences, Humanities, Social Sciences). In the context of b-learning, a constructive, optimistic, differentiated and proficient approach seems to require teachers with a highly resilient sense of personal and interpersonal awareness and openness to cultural change.

**Students’ Training (Dimension 3)**

This dimension identifies the relevance of training towards a proficient LMS usage. Considering the statistical results (eigenvalue = 8.013 and inertia = .471), there is a positive association between lack of time, the lack of interoperable systems (Linkability to other systems), the lack of technological knowledge (Students’ ICT training), and the LMS Usability.

The new dimensions of community and identity (Wenger, 1998; Zhou, 2011) that emerge in the globalization era justify new approaches in the design, implementation and development of the teaching-learning process. In fact, researchers have stressed that faculty members need more time to improve their experience in technology-based instruction, e.g., e-moderation (Salmon, 2004) and technology integration (Mishra & Koehler, 2006), with the purpose of improving technological, methodological and strategical knowledge for their own and for their students (Howell et al., 2004). From the students’ responses, it is clear that the lack of time to explore the potential of the LMS Moodle is still a major limitation. One of the interviewees assumed that: “I need more time to explore several activities and useful tools of the Moodle platform, such as chats, wikis, and forums […] or how to send assignments to the teacher! In some situations I do not know how to effectively use the platform tools and, for example, how to communicate with my colleagues” (Student11). Hence, ICT knowledge, in the context of online learning environments appears to represent an emerging need, requiring “a new set of skills for most educators and learners” (Simonson, 2005, p. 284). In an extension to this, Oh and Park (2009) argued that the lack of faculty motivation and enthusiasm to integrate technology into their distance education courses may represent the most important challenge for the implementation of effective blended teaching. Unexpectedly, the results of a study revealed that more than 36% of students surveyed consider not needing supplementary training in the use of ICT in their online courses (Kvavik & Caruso, 2005). However, according to Kenny and Pahl (2009), in an active learning approach, learning is often associated with knowledge acquisition and skills training, hence students would achieve better learning performance and higher levels of satisfaction if they are adequately trained for the effective LMS usage.

**Specific considerations**

In this research, students expressed, in general, a positive attitude towards the use of LMS Moodle, and the b-learning courses appear to ensure contextual-specific needs due to their inherent openness and flexibility. On the other hand, results point out the importance and need to create supportive environments for a more comprehensive blended structural design (e.g., based on techno-pedagogical skills) (Bates & Sangrà, 2011b; Sarirete et al., 2008), in order to harmoniously improve and sustain the learning processes and institutional contexts.

As for the methodology used in this study, the MCA seems to represent a useful technique to identify profiles and their interdependencies, and to show the structural complexity of data from a large number of categorical variables. The application of this technique allowed identifying three distinct learners’ profiles: (i) interactive learning environment-oriented, (ii) ICT teachers’ beliefs-oriented and (iii) students’ training-oriented. Once identified these profiles, one can adjust online instructional strategies in order to respond to their specific needs, bearing in mind that “engaged learning is a collaborative learning process in which the teacher and student are partners in construction knowledge” (Conrad & Donaldson, 2010, p. 8). Under this complex and multifaceted scenario, the evolution and understanding of theory and practice of truly interactive, adaptable, and co-participative b-learning environments seems to be able to guarantee interpersonal engagement, promoting and cultivating a proficient community-centered practice approach.

315
As already mentioned (Research rationale Section), in an effort going a step further, teachers’ profiles would also be incorporated in the process of enhancing the LMS environment towards more effective b-learning. Nevertheless, teaching practices on students’ learning lie in the sense of cause-and-effect relationships, with the teachers playing an active and direct role in the students’ acquisition of knowledge (Brophy & Good, 1986; Leinhardt & Smith, 1985). Consequently, teachers’ profiles should not be approached independently from the learners’ ones, but, rather, in relation to them, under the concept of teacher’s efficiency in producing desired learning outcomes to learners in a LMS-based b-learning environment, fostering the use of a variety of appropriate representational systems, examining the concept through conceptually focused and cognitively challenging tasks, and ensuring active involvement of the students within the process of knowledge construction. A research effort at this direction is within the immediate research plans of the authors.

When reflecting on the mixed method (qualitative and quantitative) adopted here, the followings could be considered. Research adopting the quantitative approach (positivism paradigm) is said to be mostly numerical and is designed to ensure objectivity, generalizability and reliability (Smith et al., 1991). One important feature of quantitative techniques is that the process of data collection is distinct from analysis. Some techniques, such as interviews or observations, nevertheless, can be interpreted either quantitatively or qualitatively. In the latter case (phenomenological paradigm), qualitative approach deals with the way people experience phenomena in the world and define its meaning. In particular, Van Maanen (1983) defines qualitative methods as an array of interpretive techniques which seek to describe, decode, translate and otherwise come to terms with the meaning, not the frequency, of certain more or less naturally occurring phenomena in the social world. They are less structured, longer and have a more flexible relationship with the respondents so the resulting data have more depth and greater richness of context (Aaker et al., 2001). It also means that the potential for new insights and perspectives is greater. Nevertheless, the data compiled by this approach may also look very ‘untidy’ because of the researcher’s lack of control on them. For instance, it is impossible for a researcher to maintain the same discussion when interviewing various individuals. This is due to the fact that humans are often encouraged to talk about unrelated things from time to time. As a result, the analysis and interpretation of the data may be very difficult. Hence, when simultaneously dealing with the problem of generalising and understanding ideas, it is better to incorporate both approaches whenever possible (mixed case). The understanding of both paradigms helps identify the ways these divergent approaches complement each other. In this perspective, the qualitative data drawn from the semi-structured interviews fed into the MCA quantitative analysis, which in turn, assisted into ‘tiding’ them up in a way that could be then interpreted accordingly in terms of identifying the three distinct students’ profiles.

From the available total number of students that used the LMS environment, 36 were selected here to define the sample space. Perhaps the interview sample is not representative of the overall group of students at the university. In practice, however, the number of required subjects usually becomes obvious as the study progresses, as new categories, themes or explanations stop emerging from the data (data saturation) (Marshall, 1996). The adopted MCA analysis has shown that this data saturation phenomenon was also observed here after analyzing 30 interviews, justifying the selection of 36 participants as an adequate starting sample size. Naturally, expansion to larger data space could further shed light upon the examined issues and research efforts towards such direction are already initiated.

From an overall perspective, the current study is a small-scale exploratory study in trying to characterize and understand dynamics patterns of a public HEI within the LMS environment. In spite of this, the present study addresses in-depth relevant issues on educational processes in LMS-based b-learning environments, in the specific context of higher education, aiming at a more educated, interactive and collaborative online community. As part of future work, we intend to scrutinize the emerging concept of Massive Open Online Courses (MOOC) recently introduced by Downes and Siemens (McAuley et al., 2010). Globally speaking, the MOOC (i.e., free online courses designed to be an all-inclusive learning experience) methodology are based on a wide blend of traditional tools, such as video lessons, evaluation tests and final exams combined with Web 2.0 tools (e.g., community of learning, wiki, blog, social media), already offered by the top institutions like Harvard, MIT or Stanford. Based upon connectivism and considering particular characteristics, such as diversity, autonomy, openness, self-organization, interactivity/connectivity for sharing knowledge, this approach can represent a unique opportunity to discover more about how, where, when, what and with whom people can learn in large open networks.
Conclusions

A new perspective of LMS within the b-learning environment has been presented here, taking into account the learners’ profiles for the identification of potential processes that act as a feedback to the LMS and could further enhance the efficiency of b-learning. Using a semi-structured interview scheme, structured in four Categories related to communication tools, collaborative and interactive networking, usage and expectations in the LMS environment, respectively, learners’ profiles from 36 students of five different b-learning courses from a HEI have been identified with the combination of qualitative and quantitative analyses applied to the acquired interview data. In particular, the MAXQDA–based content analysis of the latter resulted in a set of Subcategories per each Category, which, then, they subjected to MCA for their clustering into three different students’ profiles (interactive learning environment, ICT teachers’ beliefs & differentiation, students’ training) regarding the LMS use. This multifaceted perspective identified the need for rethinking LMS structures within the b-learning environment as a means that: (i) allows an adjustable and dynamic ecosystem that can integrate different interactive learning activities, (ii) facilitates the teachers’ ICT acquaintance to foster their intrinsic motivation, and (iii) provides students’ training strategies, so they can achieve better learning performance and higher levels of satisfaction. The concept of involving learners’ profiles in the LMS perspective provides a more pragmatic approach in the role of LMS in b-learning environment, making it more realistic according to the students’ learning needs. In this way, the students could be involved in an active and engaging learning process during the courses, giving them interesting technical and educational solutions for self-organization and learning. On the other hand, the students’ profiles contribute to the forming of original learning scenarios and more functional technical approaches. The mutual connection between LMS and the students’ profiles supports the symbiosis between students’ informal learning and the obligatory formal learning process. Based on this knowledge, in future work, the proposed approach will be extended so it would incorporate also teachers’ profiles, providing a bilateral view of the role of the active inputs of the LMS, i.e., teachers and students, towards more interactive, adaptable, and co-participative b-learning environments.

Acknowledgements

The first author would like to acknowledge the Foundation for Science and Technology (Portugal) for its financial support for this work (Grant SFRH/BD/44928/2008). Moreover, the authors would like to express their gratitude to the students of the Faculty of Human Kinetics, University of Lisbon, Portugal for their constructive discussions, opinions and comments about the main issues addressed by this research study.

References


The Effects of the E-Book System with the Reading Guidance and the Annotation Map on the Reading Performance of College Students

Liang-Yi Li, Cheng-Yu Fan, Dong-Wei Huang and Gwo-Dong Chen*

Department of Computer Science and Information Engineering, National Central University, Taiwan // lihenry12345@gmail.com // fanpaul2009@gmail.com // steven740114@gmail.com // chen@csie.ncu.edu.tw

*Corresponding author

(Submitted May 26, 2012; Revised August 16, 2012; Accepted November 5, 2012)

ABSTRACT

Novice learners demonstrate marked difficulty in using reading-study systems to read academic textbooks. One notable problem is that the process and constructed knowledge is complex. Novice learners are required to exert considerable effort in applying the process and in remembering the knowledge, resulting in lower motivation and fewer cognitive resources for reading. This study develops an e-book reading system with an integrated reading guidance module and an annotation map, and conducts an experiment for examining the effect of this system on reading, reviewing, navigational performance, and reader behavior. The results show that the annotation map significantly improves reviewing and navigational performance, but not reading performance. Log analysis identifies two problems related to reading guidance: time allocation and learner control. We discuss these results and propose recommendations for future works.

Keywords
E-Book, SQ3R, Annotation, Reading guidance

Introduction

Reading is a vital skill. Students with proficient reading skills have the potential to become enhanced self-regulated learners and, thus, demonstrate high academic achievement. However, most students are not proficient in reading. Many students even lack the basic reading skills necessary to perform future job-related tasks (Artis, 2008), which will greatly affect their future study and work.

Several reading-study strategy systems provide clear guidelines to help students learn and practise techniques that imitate the behaviors of highly proficient readers (e.g., SQ3R, 3R, and KWL; Al-Khateeb & Idrees, 2010; Artis, 2008; McDaniel, Howard, & Einstein, 2009; Robinson, 1970). SQ3R, which consists of five steps (surveying, questioning, reading, reciting, and reviewing), is the most popular reading-study system and was primarily designed for expository text, particularly academic textbooks.

Although SQ3R is a purposeful and meaningful reading method in which students practise different reading strategies, it is cumbersome for novice learners to learn and use (Flippo & Caverly, 2000; Huber, 2004). The SQ3R process is complex, and the knowledge constructed during this process is comprehensive and varied. Novice learners must expend more cognitive and behavioral effort in operating and managing the process and knowledge before they become experienced. Their effort may impede reading comprehension when learners are unfamiliar with this method. This impediment may lower their motivation for using and practising this strategy (Artis, 2008).

Because the SQ3R process is complex, novice learners must use more cognitive resources to remember the SQ3R steps, when to use these steps, what purpose each step has, and how to perform each step. Although training and practice can facilitate the learning process, it consumes time and effort. Teachers typically address this method for a brief period and ask students to use it by themselves after class.

The learning products that are constructed and acquired during the SQ3R process are comprehensive and varied: section titles, keywords, an overview in the surveying step, questions in the questioning step, comments and crucial points in the reading step, and summaries in the reciting step. The relationship among the products forms a hierarchical structure. For example, when creating a question for a section title, several pieces of highlighted text and comments are generated to answer the question or explain a keyword, and a summary is written for the section. The hierarchical structure is useful for remembering and reviewing the learning products. However, they are distributed among different pages, and students have difficulty perceiving and remembering the structure and relationship.
This study designed an e-book reader that integrates SQ3R reading strategies with two scaffolding tools, a reading guidance module, and an annotation map, to solve the previously mentioned problems. The reading guidance module reminds readers of the purpose and task of each SQ3R step and provides examples for imitation. The annotation map, which integrates the annotations distributed among different pages into a hierarchical structure, is designed to support reading, reviewing, and navigating. In addition, we conducted an experiment to determine the effect of this e-book reader on the reading behavior and performance of readers.

Related works

Reading strategies and SQ3R

Reading strategies, which play a vital role in reading comprehension, have been recognized as effective approaches in increasing reading comprehension (Huang, Chern, & Lin, 2009). Successful readers apply several reading strategies for reading comprehension, such as recognizing text structure, posing questions, reflecting on behavior or the process, monitoring comprehension, organizing graphs, taking notes, and rereading (Sung, Chang, & Huang, 2008; Yang, 2006). Teaching every student to master these strategies in general courses is challenging. Therefore, several methods that bundle a set of reading strategies provide easily applied guidelines for students to practise independently during their reading processes.

SQ3R is the most popular reading method. Robinson (1970) developed this method by incorporating several higher-level study strategies consisting of five steps: surveying, questioning, reading, reciting, and reviewing. The surveying step involves skimming section headings and subheadings to identify chapter content, thus assisting readers in understanding the structure of the chapter. In the questioning step, readers ask questions based on headings and subheadings and transform the headings into questions to guide their reading. The reading step involves reading the text to answer the questions. In the reciting step, readers recall the answers and crucial points in the chapter. In the fifth step, readers reflect on the original questions and answers, organize relevant information, and review the information repeatedly.

Previous studies have investigated the effects of SQ3R on reading comprehension. However, the sparse evidence of the efficiency of SQ3R is inconsistent (Baier, 2011; Cantu, 2006; Carlston, 2011; McCormick & Cooper, 1991). For example, Baier (2011) observed that SQ3R significantly improved the overall comprehension scores of fifth grade students reading expository texts. However, Cantu (2006) found that the SQ3R learning method demonstrated little or no improvement over traditional learning methods. Such inconsistency may result from the varying material, instruction, task demand, or student variables (Flippo & Caverly, 2000).

Although insufficient evidence exists to prove that the SQ3R system is an effective reading strategy, it remains the most popular method among teachers. SQ3R imitates the reading patterns of proficient readers, provides useful examples for poor readers to improve, and gives meaning and purpose to reading. SQ3R techniques, such as questioning and summarizing, have also been separately proven to enhance reading comprehension and may foster in students a familiarity with the techniques.

Scaffolding

Scaffolding is a process in which assistance is provided to students when needed, and gradually reduced when their competence increases (Molenaar, Roda, Boxtel, & Sleeegers, 2012). There are two types of scaffolding: static and dynamic. Static scaffolding is the same for all students and remains constant over time, whereas dynamic scaffolding involves analyzing student behavior to provide the most appropriate assistance. Previous studies have compared the advantages among dynamic, static, and no scaffolding, revealing that dynamic scaffolding yields greater benefits than does static scaffolding; however, static scaffolding provides more advantages than does no scaffolding (Azevedo, 2005; Azevedo, Cromley, & Seibert, 2004; Moos & Azevedo, 2008). Although dynamic scaffolding provides more advantages than static scaffolding does, dynamic scaffolding is difficult for teachers to implement and has been successfully applied only in structured domains (Li & Chen, 2009). In ill-structured domains, developing student models is difficult.
Previous studies have developed different scaffolding tools for self-regulated learning; however, few studies have developed tools to support SQ3R. Several studies have developed useful systems to support reading-study systems. For example, Kozminsky (2010) designed a system that provides several graph organizers for students to organize learning products during the SQ3R reading process. Zhang, Cheng, Huang, and He (2002) designed a distance-learning support system that helps students perform the SQ3R process by providing various tools such as a material map to assist students in browsing as well as memo tools to help students ask questions. However, these studies have not evaluated their levels of effectiveness.

Annotation

Annotation is a useful technique for adding information to existing documents and allows users to extend the meanings and contexts of textbooks (Glover, Xu, & Hardaker, 2007). Annotation consists of two types: marking (i.e., underlining or highlighting) portions of text and writing notes in margins or between lines (Ovsiiannikov, Arbib, & McNeill, 1999). Marking text, which identifies key parts of the document, is the most frequently used reading aid (Ovsiiannikov, Arbib, & McNeill, 1999) and serves as three functions to support text reading. First, marking text is an encoding process that identifies which parts are relevant and which parts are not. Second, marked text plays the role of visual signaling and can guide attention and reduce unnecessary visual searching when readers revisit the text. Third, marked text can be seen as a contextual cue associated with the text in learning materials. Contextual cues can help readers to recall the content and context of the marked text.

With the rapid growth in the number of digital learning materials, numerous annotation systems have been developed to support annotation by using PCs, the Web, and mobile phones (Glover, et al., 2007; Hoff, Wehling, & Rothkugel, 2009; Ovsiiannikov, Arbib, & McNeill, 1999; Rau, Chen, & Chin, 2004). Except for basic annotation functions (e.g., highlighting text and adding notes), several annotation systems also support shared annotation (Wolfe, 2008), multimedia annotation (Hwang, Wang, & Sharples, 2007), and annotation management (Li & Chen, 2010). Although these annotation systems provide comprehensive and varied functionalities to support reading, they are not integrated with reading strategies, particularly the SQ3R strategy.

System design and implementation

To solve the identified problems, we developed an e-book system. The e-book reader interface is divided into three sections: the e-book, annotation map, and guidance section (Fig. 1). The e-book section presents the e-book content. The annotation map section allows students to develop and organize annotations into a hierarchical structure. The guidance section presents the SQ3R steps in sequential buttons.

Three modules were designed for this system: annotation, annotation map, and reading guidance. The annotation module provides basic annotation functions consisting of underlining, highlighting, and commenting. The annotation map module supports students in creating and organizing annotations.

Five types of tree nodes can be created in the annotation map section: heading nodes, which represent the underlined section titles and keywords in the surveying step; questioning nodes, which represent notes associated with the questions created for section titles in the questioning step; highlighting nodes, which represent the highlighted text segments created in the reading step; commenting nodes, which represent the notes made during the reading step; and reciting nodes, which represent the notes associated with summaries created for the sections in the reciting step. The node icons are represented with their respective annotation functions for students to identify the role of each node easily.

Highlighting or underlining a piece of information and organizing it into a hierarchical structure is simple. When marking a within-page text segment, a student can use a PC mouse to right-click a node in the annotation map section to access a context menu. The student can click the Heading or Highlight menu item and immediately underline or highlight this marked text segment. A heading node or highlighting node with a title assigned by students is generated under this clicked node. The student can also click the Question, Comment, or Recite menu item to generate a questioning, commenting, or reciting node.
The reading guidance module guides students to read by using five buttons: survey, question, read, recite, and review (Fig. 1). Before a student reads a chapter, the survey button can be activated. When the student presses the survey button, a message informs students of the purpose of this step and shows how to perform this step. When turning to a page that includes a section title or keyword, a student can mark it and click the Heading button presented above the annotation map section or the Heading menu item on the context menu, which is raised when right-clicking an existing heading node. The marked text is then immediately underlined, and a heading node representing the underlined section title or keyword is immediately added to the first level of the hierarchical structure or below the clicked heading node. After completing this step, the student can press the question button to enter the questioning step.

In the questioning step, a student generates questions for the section/subsection titles or keywords. The student can click the Question menu item on a heading node and a question form is presented for the student to enter the title and description of the question. A question node representing the question is created below the heading node, and a question picture associated with the question node is added to the page on which the section title exists.

When pressing the reading button, the student can read the pages in the first section and click the Highlight menu item to highlight a text segment or Comment to take notes. A highlighting node with a title assigned by the student or a commenting node is generated under the node in which the context menu was raised. After completing the reading step, the student can press the recite button. In the reciting step, students create a reciting node. Creating a reciting and commenting node is the same as creating a questioning node, and creating a highlighting node is the same as creating a heading node. Finally, the student can use the annotation map to review the content by double-clicking a node to go to the corresponding page.

This e-book system provides several types of static scaffolding to guide students during the SQ3R reading process. Table 1 lists these types of scaffolding.
Table 1. Scaffolding strategies used in the system

<table>
<thead>
<tr>
<th>Scaffolding</th>
<th>Module</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructing</td>
<td>Reading guidance</td>
<td>The descriptions in each step explain the purposes and instruct students in how to perform the task.</td>
</tr>
<tr>
<td>Modeling</td>
<td>Reading guidance</td>
<td>The SQ3R process models how a proficient reader reads a textbook. The examples provided in each step teach readers how to conduct each SQ3R step.</td>
</tr>
<tr>
<td>Cognitive structure</td>
<td>Reading guidance and annotation map</td>
<td>The full process and current step is presented in the guidance section, which guides readers to read in a disciplined manner. The annotation map facilitates visualizing the products of the SQ3R knowledge construction process.</td>
</tr>
</tbody>
</table>

Method

The proposed e-book reader primarily involves two modules: reading guidance and an annotation map to support the SQ3R reading process. The reading guidance module, which consists of the scaffolds of instructing, modeling, and cognitive structure, reminds readers of the steps and purposes of SQ3R and demonstrates examples to help readers perform these steps. Therefore, the first question we would like to examine is whether the reading guidance module can effectively guide readers complete the SQ3R process.

The annotation map organizes the annotations created when using SQ3R in a hierarchical structure. Organizing learning products is a generative process that helps students effectively retain and recall information from long-term memory. The annotation map also reminds readers which annotations were created, including information regarding their relationships. When a reader must compare or synthesize information between pages, they can refer to this map without navigating between pages. We anticipated that the annotation map would be able to help readers remember and understand the content of particular textbooks. Therefore, the second question we sought to answer is whether the annotation map improves reading, reviewing, and navigational performance.

To answer the two questions, a quasi-experiment was conducted. In addition, we also executed a separate study in which a questionnaire and an unstructured interview were conducted to understand thoroughly student perceptions and levels of satisfaction regarding the use of such a system.

Study one: The quasi-experiment

Design

To answer Questions 2, a quasi-experiment was conducted. The between-subjects factor was the e-book reader system (i.e., the systems with reading guidance [S1] and with reading guidance and annotation map [S2]). Five dependent variables were used to examine annotation map effectiveness on reading and navigational performance: the score for reading comprehension after reading, the score for reading comprehension after reviewing, the time spent for navigational tasks, the number of pages visited for completing navigational tasks, and the number of navigational tasks completed in two attempts. We also used system logs to answer Question 1.

Participants

Forty-one first-year computer science undergraduate students recruited from National Central University in Taiwan participated in this experiment. Each participant was paid US$10 for participating. All of the participants had studied the Fundamental Computer Science course for two months. A total of 21 participants were in the S1 group and 20 participants were in the S2 group. Because prior knowledge can affect learning achievement, the participants were stratified into five levels, based on their mid-term exam scores in this course (the mid-term exam was conducted one week before the experiment) and were then randomly assigned to two groups based on the strata. No significant effect on mid-term exam scores ($t = -.259$, $p = .797 > .05$) was observed between the S1 group (Mean = 86.76, SD =
and the S2 group (Mean = 87.60, SD = 11.04). Therefore, the two groups had the same level of prior knowledge.

Materials

The materials consisted of two reading systems (S1 and S2), two reading materials, two comprehension test questionnaires, and a paper handout. The language used in these materials was Mandarin Chinese.

The S1 group used our system without the annotation map section. The annotation functions for the respective steps were provided in a toolbar above the e-book section. The S2 group used our system.

The reading materials were extracted from a textbook from Fundamental Computer Science. Chapter Three, “Data Storage,” was selected as practice material. Chapter Five, “Computer Organization,” was selected as the experimental material and included 24 pages. We extracted 17 pages, which could be finished within 60 minutes.

The paper handout consisted of nine task descriptions (Tasks 1–9). Each description in Tasks 3, 6, and 9 was a picture (which was extracted from the reading material) for the participant to locate. Each description in Tasks 1, 4, and 8 was a paragraph extracted from a specific web page that participants had to locate. Each description in Tasks 2, 5, and 7 was a keyword for which the participants had to locate the definition.

Two comprehension tests, one for reading and one for reviewing, were used to evaluate reading performance. The comprehension test included ten open-ended questions for reading, consisting of four remembering, four comprehension, and two application questions. The comprehension test for reviewing comprised four memory, four comprehension, and two application questions. These questions were selected from textbook exercises. The scores of the two comprehension tests were evaluated by a graduate student and a postdoctoral researcher who had majored in computer science. After measuring the scores of each student, they had a meeting to reach a consensus.

Procedure

The experiment was conducted in a usability laboratory and was divided into three phases: instructing, reading, and reviewing. All activities in the reading systems were logged in Microsoft Access database. In the instructing phrase, the entire experimental procedure was first explained to the participants. The participants were then instructed in the SQ3R reading strategy for 30 min. Finally, they were asked to read the practice material and use the assigned system without any time constraints.

For the reading phase, the participants were allowed 60 minutes to read through the experimental material and use the assigned reading system, after which, they were administered a reading comprehension test.

The reviewing phase was conducted two weeks after the reading phase. Participants were given 15 minutes to review the material while using their assigned system. All annotations were preserved in the systems. After fifteen minutes they received the paper handout and completed a comprehension test and navigational tasks.

Result of the experiment

Reading behaviors during the SQ3R process

The first analysis was used to examine how participants allocated time for each SQ3R step through descriptive statistics (Table 2). This allocation of time seems reasonable. However, except for the reading step, the standard deviations of these steps are large, which may indicate that time allocation for the steps differed significantly among the participants. Therefore, a cluster analysis of the time percentages spent in each of the five steps was conducted. This analysis classified the time allocation patterns into a homogeneous group. The K-mean method was used.
As shown in Table 3, four clusters were identified. The participants in Cluster 1 spent more time in the surveying and questioning steps, indicating them as slow SQ3R readers. However, because less time remained for the reciting and reviewing steps, they skipped these steps. By contrast, the participants in Cluster 2 demonstrated themselves as fast SQ3R readers who quickly completed the survey and question steps with more time remaining for the reciting and reviewing steps. The participants in Cluster 3 invested most of their time in the reading step and quickly passed the other steps, whereas the participants in Cluster 4 invested little time in the reading step.

Reading performance

An independent sample t-test was conducted to compare the comprehension scores of the two groups after reading. The results showed no significant effects of the reading systems on comprehension scores after reading between the S1 (Mean = 55.81, SD = 24.72) and S2 groups (Mean = 60.05, SD = 20.23; t = −1.988, p = .054), indicating that the participants in both groups had the same reading performance.

Navigational performance

Navigational performance is evaluated using three variables: time spent, navigational path, and the number of tasks completed within two tries. The number of nodes a subject visited for completing the navigational tasks was counted as one navigational path.

Three independent sample t-tests were conducted to examine the effect of the reading systems on navigational performance. The results revealed a marginally significant effect of reading systems on time spent (t = −1.988, p = .054), indicating that the participants using S1 (Mean = 258.14, SD = 107.57) spent less time than did the participants using S2 (Mean = 323.55, SD = 102.86). Additionally, there was a significant effect of reading systems
on the number of tasks in which the information was found within two tries $(t = -2.38, p = .022 < .05)$. The result indicated that the participants using S2 (Mean = 4.70, SD = 2.70) completed significantly more tasks within two tries than did the participants using S1 (Mean = 3.10, SD = 1.48). However, the navigational path was not significant $(t = .726, p = .472)$.

**Table 5. Navigational performance between the systems with and without annotation map**

<table>
<thead>
<tr>
<th></th>
<th>Without annotation map</th>
<th>With annotation map</th>
<th>$t$-statistic</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time spent (second)</td>
<td>Mean = 258.14</td>
<td>Mean = 323.55</td>
<td>-1.988</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>SD = 107.57</td>
<td>SD = 102.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Navigational path</td>
<td>Mean = 52.10</td>
<td>Mean = 46.50</td>
<td>0.726</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>SD = 20.09</td>
<td>SD = 28.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of tasks</td>
<td>Mean = 3.10</td>
<td>Mean = 4.70</td>
<td>-2.38 *</td>
<td>39</td>
</tr>
<tr>
<td>completed within two</td>
<td>SD = 1.48</td>
<td>SD = 2.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tries</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p*-value < 0.05, ** *p*-value < .01

**Reviewing performance**

Because reading performance may affect reviewing performance, a one-way analysis of covariance (ANCOVA) was used to examine the effect of the comprehension scores of the two groups after reviewing. The covariate was their comprehension score after reading. A significant effect of the reading systems on comprehension scores after reviewing was found ($F = 5.218, p = .028 < .05$), indicating that the participants using S2 (Mean = 62.21, SD = 2.86) had significantly higher comprehension scores than did the participants using S1 (Mean = 53.08, SD = 2.79) after reviewing.

**Study two**

Study two involved administering a questionnaire and conducting an unstructured interview to collect the qualitative data for understanding student perceptions and levels of satisfaction regarding the use of such a system. All students from the S2 group in Study One were selected to participate in this study, which continued for two days. On the first day, the experimenters instructed the participants in how to use this e-book system. The participants then read a text consisting of 13 pages. Finally, the participants were administered an exam. On the second day, the participants read a text consisting of 16 pages. The participants then completed a questionnaire. Finally, the experimenters randomly selected three participants and interviewed them individually.

**Findings of study two**

**Table 6. Satisfaction questionnaire**

<table>
<thead>
<tr>
<th>Items</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading guidance can help me learn SQ3R efficiently.</td>
<td>4.05</td>
<td>0.51</td>
</tr>
<tr>
<td>I can understand what purpose each step has and how to perform each</td>
<td>4.10</td>
<td>0.45</td>
</tr>
<tr>
<td>step.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I think the instructions are clear.</td>
<td>4.20</td>
<td>0.41</td>
</tr>
<tr>
<td>I have followed the instructions for each step to apply SQ3R.</td>
<td>3.70</td>
<td>0.80</td>
</tr>
<tr>
<td>I can demonstrate high reading performance if I follow the</td>
<td>3.75</td>
<td>0.79</td>
</tr>
<tr>
<td>instructions.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The annotation map can help me to remember the structure of reading</td>
<td>4.55</td>
<td>0.51</td>
</tr>
<tr>
<td>materials.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The annotation map can help me to recall the content of reading</td>
<td>4.30</td>
<td>0.66</td>
</tr>
<tr>
<td>materials.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The annotation map can improve my reading performance.</td>
<td>4.30</td>
<td>0.57</td>
</tr>
<tr>
<td>The annotation map can help me to review the learning materials</td>
<td>4.20</td>
<td>0.77</td>
</tr>
<tr>
<td>efficiently and effectively.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The annotation map can help me find a specific annotation easily.</td>
<td>4.30</td>
<td>0.80</td>
</tr>
<tr>
<td>The annotation map can help me find a specific paragraph easily.</td>
<td>4.47</td>
<td>0.61</td>
</tr>
<tr>
<td>I think that using the e-book system to read textbooks can enhance</td>
<td>3.85</td>
<td>0.59</td>
</tr>
<tr>
<td>my reading performance.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I think that the e-book system is easy to learn and use.</td>
<td>4.15</td>
<td>0.59</td>
</tr>
</tbody>
</table>
Table 6 lists the questionnaire results. The value of Cronbach’s alpha for the 13 items was 0.84, indicating a highly acceptable level of reliability. The results showed that the participants had a positive attitude regarding our system. Additionally, we summarized several focal comments and suggestions proposed by the three participants in the unstructured interviews. The summary is listed as follows:

- The SQ3R strategy is cumbersome (n = 3).
- I am not used to asking questions about the section title before reading (n = 3).
- The number of words in the descriptions of the five SQ3R steps is too many. Although the descriptions are very clear, I suggest that the number of words be shortened (n = 3).
- I can remember the instructions for each step after my first reading. I believe that showing the descriptions is not necessary in subsequent reading (n = 3).
- I sometimes inadvertently view the annotation map while reading. The map distracted my attention (n = 1).
- Organizing annotations hierarchically is inconvenient. I must interrupt longer periods of reading when organizing annotations into the hierarchical structure (n = 3).
- When I review the reading materials, I first view the annotation map, then sequentially browse the materials in the e-book (n = 3).

Discussion

This study developed an e-book system to facilitate textbook reading through the SQ3R process. A quasi-experiment was designed to answer two questions. For Question 1, the descriptive statistics and clustering analyses revealed two problems. First, the time allocated among the participants showed a large difference. It shows that the participants had problems monitoring their reading process and allocating their time. Second, the readers were unwilling or did not have the ability to question and summarize textbook content. These results indicated that the current status of this reading guidance application does not effectively guide readers to complete the SQ3R process.

To overcome these problems, we offer several additional design suggestions. In this study, reading guidance provided novice learners with full control. However, findings from previous research suggest that providing novices with full control imposes too much of an extraneous load (Ayres & van Gog, 2009; Corbalan, Kester, & van Merrienboer, 2008). We suggest that this reading guidance module should increase system control to regulate novice learners before they can implement the SQ3R process independently. Second, solely providing instruction in how to conduct the SQ3R steps is insufficient; the system should provide effective scaffolds to support readers in performing highly cognitively demanding tasks (e.g., by explicitly asking readers to write crucial points or answer questions in the reciting step). Third, this system should support readers in monitoring and maintaining awareness of their reading progress, for example, by providing time management functions to raise awareness in novice readers of the time allocated and used.

Regarding Question 2, the participants did not significantly differ in comprehension scores after the reading phase. Based on student questionnaire results and comments, the annotation map supports readers in remembering the structure and content of textbooks. However, it also has several drawbacks that impede reading comprehension. First, the annotation map may draw readers’ attention away from reading. The distraction effect was also found in previous studies related to structure overview in hypertext reading (Hofman & Oostendorp, 1999). Second, interruptions in reading or attention shifts can impair learning and retention (DeStefano & LeFevre, 2007). Because users who employ the e-book reader system with the annotation map must additionally select a node they wish to highlight or comment on, the interrupted time for them is longer.

The results for reviewing performance revealed that participants using S2 had significantly higher comprehension scores than did participants using S1 after reviewing. This is because the annotation map reminds readers of what has been constructed and performed and helps them navigate between pages to glean crucial points quickly. Therefore, the participants with the annotation map in this study received higher comprehension scores when they had a limited time for review.

The S2 group completed significantly more tasks in two attempts, as compared to the S1 group. However, the S2 group spent a marginally significantly longer time to complete the tasks than did the S1 group. The navigational path
between both groups did not yield a significant difference. The following reasons might explain these results. First, within-text annotations are contextual cues associated with textbook content. The annotation map presenting an overview of the annotations helps readers recall the associations. If readers can correctly recall which annotation is associated with the target, then they can directly link to the target page with one click of a mouse on the annotation node. Therefore, S2 enhanced the performance of finding targets in two attempts. Second, when participants cannot clearly remember the location of the target, S1 users sequentially browse the pages through the links provided by the materials. By contrast, S2 users first scan the hierarchical structure to guess the nodes that may be associated with the target and then link to these pages associated with the nodes with one click. Because the S2 users in this study spent time thinking, they used more time than did the S1 users. Third, although the S1 users spent less time finding the targets, they still negotiated a longer navigational path. This may be because S1 users sequentially browse pages through the previous and next links provided by the materials. This navigational strategy allows quick switching between pages. Therefore, the length of the navigational path of the S1 users in this study was longer than S2, although it is not statistically significant. These results indicate that the annotation map supports students in accurately locating targets.

The results related to the annotation map show that the annotation map is helpful for reviewing and navigating activities. However, reading performance was not significantly improved. We believe that this may be due to the effects of distraction and interruption. To solve these problems, we suggest that teachers change the time for organizing annotations from during reading to after reading. Accordingly, students need only highlight text and add notes during their reading, and organize the annotations after the reading step. Thus, the two effects of distraction and interruption should disappear.

Several previous studies that have asked students to organize concept maps during reading have shown that concept maps can improve reading performance (Kwon & Cifuentes, 2009). However, these results are inconsistent with those of our study. An annotation map is a hierarchical concept map; the relationship between the nodes is top-down. The concept maps constructed in previous studies have been network structures, on the other hand, so the relationship between the nodes has been more complex. Regarding this complexity, students who construct network concept maps engage in a deeper encoding process than do students who construct hierarchical concept maps. Engaging in a deeper encoding process may result in higher reading performance. However, hierarchical concept maps are easily organized, and constructing network concept maps requires substantially more effort. Although constructing network concept maps may result in higher reading performance, the effort may also lower students’ motivation to use it. Determining which concept map is more feasible in practical classes should be investigated further.

The proposed e-book system was designed to support student reading during the SQ3R process. However, the approaches and strategies students use during their usual reading processes vary. SQ3R is not suitable for all students. We believe that SQ3R can benefit students who do not apply consistent reading strategies or do not demonstrate proficient reading skills. Therefore, teachers must determine students’ reading abilities and assign adaptive levels of system control for each student before introducing this strategy and system. For proficient readers, this system can provide the annotation tools and the annotation map. Students can read textbooks by using their usual reading processes and strategies without requiring reading guidance. For poor readers, teachers can assign the e-book system with reading guidance. When students have more experience in using the SQ3R strategy, teachers can reduce the level of system control. For example, teachers can cancel the instruction in each step when a student has used the system for once or twice, or cancel the questioning step when a student demonstrates the ability to ask questions in the questioning step. By applying this fading process, poor readers have the opportunity to become proficient readers.

Training students in reading strategies consumes considerable time and effort. In Taiwan, teachers do not have adequate time to accomplish this. Therefore, applying this system can solve this problem. Teachers can incorporate this system into a specific course learning activity, such as a homework assignment. Teachers can assign students reading based on their teaching progress. Students can then use the system to complete their reading assignments. Because the system is combined with a learning activity, students should have more motivation to use it. Thus, teachers would not require too much additional course time to instruct students. In addition, students (particularly Taiwanese college students, who lack regular reading habits and sometimes demonstrate low reading skills) can learn diverse reading strategies and regulate their reading effectively.
Conclusion

The proposed e-book system with the reading guidance module and annotation map was anticipated to support the SQ3R reading process for novice learners. An experiment revealed several findings and prompted suggestions. First, the annotation map is a hierarchical concept map that can improve reviewing and navigational performance, but not reading performance. Reading performance may be impeded by the effects of interruption and distraction. Second, the results of reading performance are inconsistent with those of previous studies, possibly because previous studies have used network concept maps, whereas the annotation map is a hierarchical concept map. This prompted two questions: (1) Do students prefer hierarchical or network concept maps? (2) Does organizing a network concept map after reading result in higher reading and reviewing performance than does organizing a hierarchical concept map after reading? In addition, SQ3R is not the usual way for students to read textbooks. When students have limited experience in SQ3R, they encounter difficulties in monitoring their reading progress and performing specific reading strategies that are highly cognitively demanding. This study determined that the types of reading guidance that provide only the static scaffolding of instructing, modeling, and cognitive structure is insufficient. We offer additional design suggestions consisting of offering more system control for novice learners and providing scaffolding tools to support readers in performing tasks that are cognitively demanding, as well as to monitor their reading progress. In the future, we will redesign the system based on the suggestions and will incorporate it in a practical course for a longer duration.

Acknowledgements

This project was supported by the National Science Council of Taiwan under contract numbers NSC102-2511-S-008-017-MY2 and NSC 100-2511-S-008-003-MY3.

References


Potential Negative Effects of Mobile Learning on Students’ Learning Achievement and Cognitive Load—A Format Assessment Perspective

Hui-Chun Chu
Department of Computer Science and Information Management, Soochow University, 56, Sec. 1, Kui-Yang St., Taipei 100, Taiwan // carolhcchu@gmail.com

(Submitted June 12, 2012; Revised March 27, 2013; Accepted May 11, 2013)

ABSTRACT

Owing to the advancement of mobile and wireless communication technologies, an increasing number of mobile learning studies have been conducted in recent years. In a mobile learning environment, students are able to learn indoors and outdoors with access to online resources at any time. However, whether or not new learning scenarios that combine both real-world contexts and digital-world resources are beneficial to the students has been questioned. Moreover, it is also interesting to probe whether the existing e-learning strategies are effective when situated in those mobile learning scenarios. In this study, an in-field activity on an indigenous culture course of an elementary school with a formative assessment-based learning strategy was conducted to investigate the possible negative effects of mobile learning by analyzing the students’ cognitive load and learning achievement. It is interesting to find that, without proper treatment, the performance of students using those existing online learning strategies, known to be “effective,” might be disappointing or may even negatively affect the students’ learning achievements. Furthermore, the negative effects could be due to the heavy cognitive load caused by an improper learning design. Such findings offer good references for those who intend to design and conduct mobile learning activities.

Keywords
Mobile learning, Ubiquitous learning, Formative assessment, Cognitive load, Technology-enhanced learning

Background and objectives

In the past decade, the advancement of mobile technology has offered an opportunity to provide learning supports at anytime and anywhere. Various studies have reported the benefits of applying mobile technologies (e.g., personal digital assistants, smart phones, or portable computers) to the learning activities of various courses, including science, social science, and language courses (Hoppe, Joiner, Milrad, & Sharples, 2003; Liu, Wang, Liang, Chan, Ko, & Yang, 2003; Hsu, Hwang, & Chang, 2010; Sharples, Taylor, & Vavoula, 2007; Wong & Looi, 2011; Wong, Chin, Tan, & Liu, 2010; Zhang et al., 2010). For example, Chen, Kao, and Sheu (2003), based on scaffolding theory, developed a bird-watching mobile learning system that enables learners to observe birds outdoors, while gaining information about the birds via mobile devices. Moreover, teachers can send questions to students by means of these devices. Later, Chen, Chang, and Wang (2008) also employed the scaffolding theory to conduct mobile learning activities. In the meantime, Chu, Hwang, Huang, and Wu (2008) conducted several outdoor learning activities in a butterfly ecology garden by integrating mobile learning environments with electronic library facilities. Recently, Hwang, Wu and Ke (2011) proposed an interactive concept-mapping approach for conducting mobile learning activities for natural science courses.

During the process of mobile learning, students interact with authentic contexts through words, pictures, sounds, animations, and images, all provided by the mobile devices. Compared with traditional instruction or conventional web-based learning, mobile or ubiquitous learning scenarios could be much more complex for learners since they need to simultaneously deal with learning materials in both the real world and the digital world. In the past decade, many studies have shown that proper mobile learning strategies or tools need to be considered to help the students acquire the expected learning achievements in real-world environments (Chu, Hwang, & Tsai, 2010; Chen, Chang, & Wang, 2008; Chen, Kao, & Sheu, 2003; Liu, Lin, Tsai, & Paas, 2012; Looi, Zhang, So, Chen & Wong, 2010). Most of those studies have focused on whether the students’ learning performance can be improved by the mobile learning system or by using proper traditional instruction strategies (Hwang, Wu, Tseng, & Huang, 2011; Wu, Hwang, Su, & Huang, 2012). Only a few studies have been conducted to investigate the impacts of mobile learning on students’ cognitive load (e.g., Liu et al., 2012), which might be an important indicator to evaluate the effectiveness of applying well-recognized web-based learning strategies to mobile learning activities (Ozcinar, 2009).
Cognitive load theory was proposed in the 1980s. It is concerned with the way in which humans’ cognitive architecture deals with learning objects during the learning process or when performing a particular task. Human cognitive architecture is composed of working short-term and long-term memory, in which all conscious cognitive processing occurs; moreover, cognitive load has been recognized as being closely related to the demand of working memory resources during the learning process (Paas, Renkl, & Sweller, 2003). Baddeley and Hitch (1974) found that human working memory can handle only a very limited number of novel interacting elements, possibly no more than two or three; that is, improper design of learning content and instructional strategies is likely to increase the cognitive load of students owing to overloading their working memory.

In this study, the students’ cognitive loads as well as their learning outcomes of an in-field mobile learning activity are investigated. In-field mobile learning refers to the learning activities conducted in the field using mobile devices, such as learning activities in museums, ecology parks or historical buildings. The students can benefit as a result of the activities promoting their interest in and motivation regarding the learning subjects when they learn in the physical field rather than learning in the classroom. In an in-field mobile learning activity, students need to observe and interact with the real-world learning targets while at the same time paying attention to the learning guidance or supplementary content provided by the learning system via the mobile device and wireless communications. Consequently, it is likely that the learning burden of individual students will exceed what they are able to cope with, implying that measuring their cognitive load as well as their learning achievement is necessary.

In a mobile or ubiquitous learning activity, the learning materials from both the real-world and the digital-world environments could be a noticeable source of mental load, since the amount and complexity of the materials are likely to increase interactions among the task, the students, and the learning materials (Paas & van Merriënboer, 1994; van Merriënboer & Sweller, 2005; Wirth, Kunsting, & Leutner, 2009); moreover, the use of improper learning strategies could increase students’ mental effort as well. Therefore, it is worth investigating both the effectiveness and the cognitive load when conducting learning activities with those well-known learning strategies.

Consequently, it has become an important and challenging issue to re-examine whether those effective web-based tutoring or assessment strategies can benefit mobile or ubiquitous learning, in particular, the formative assessment strategy that has been recognized as being one of the most effective methods for web-based learning (Chen & Chen, 2009; Heinrich, Milne, & Moore, 2009; Hung, Lin, & Hwang, 2010; Wang, 2008, 2011). When learning with the formative assessment strategy in the physical mobile environment, students were asked to observe real-world learning targets to answer a set of questions. The characteristic of this approach is that the learning system does not provide correct answers to the students when they fail to correctly answer some questions; instead, only some clues are provided to help them find the answers by themselves. That is, for those students who are inexperienced in learning with such an approach or whose learning progress is not as good as expected, their cognitive loads may be overloaded since they need to look for answers in the physical world and answer the questions repeatedly until the learning system confirms that they give the correct answer to each question three times. Paas and van Merriënboer (1994) have indicated that learners’ cognitive load may increase owing to the characteristics of the task, the subject performing the task, or the interactions between both. Kalyuga (2009) has further indicated that learning strategies aiming to increase germane cognitive load (the cognitive load that helps students learn in a more effective way) might lead to intrinsic cognitive load (the cognitive load that might seriously affect the students’ learning performance) owing to the complexity and difficulty of the learning materials and the knowledge level of the learners. In the in-field mobile learning environment with the formative assessment approach, students’ cognitive load may be high owing to the pressure caused by the limited learning time, the number of questions to be answered, students’ lack of experience in learning with mobile devices, and the need to repeatedly answer the answers until all of the criteria of the learning task are fulfilled. That is, it is quite possible that students have high cognitive load owing to the nature of the formative assessment strategy and the complexity of the in-field mobile learning environment that combines the physical learning materials and the digital learning resources. This is especially so for novices. Therefore, it is important to investigate the potential negative effects of such an in-field mobile learning approach.

This study aims to investigate the effect of applying web-based learning strategies that are known to be effective in mobile learning activities. To achieve this objective, a formative assessment-based learning system was developed for use in a mobile learning activity. Such learning strategies have been reported as being effective in web-based learning environments (Wang et al., 2004; Wang, 2008, 2011). Moreover, the learning achievements and cognitive load of the students who participated in the learning activity were measured to evaluate the effects of this approach.
Literature review

Cognitive load

Cognitive load theory is a multidimensional construct that represents the different loads that performing a particular task imposes on the cognitive system (Paas, Tuovinen, Tabbers, & van Gerven, 2003; Sweller, van Merriënboer, & Paas, 1998). Scholars have reported that both the background of the learner and the environmental context could be sources of cognitive load (Paas & van Merriënboer, 1994); moreover, learning difficulty is highly relevant to the increase in cognitive load (Sweller, 1988; Kalyuga, 2009). Researchers have pointed out the existence of three types of cognitive load (Sweller, van Merriënboer, & Paas, 1998):

- “Intrinsic cognitive load” refers to the inherent structure and complexity of the instructional materials. It increases when the element of interactivity is added. The element of interactivity depends on the number of different types of information that learners must integrate and absorb at the same time. On the other hand, “intrinsic cognitive load” is relevant to how much information the working memory needs to deal with at the same time.

- “Extraneous cognitive load” or “ineffective load” refers to the degree to which a task influences learning. It is related to the poor design of the instructional strategy, which does not take the necessary instructional variables into account. The worse the instructional design, the more difficult the learning task will become.

- “Germane cognitive load” or “effective load” refers to an instructional design that facilitates the learning process by properly taking the necessary instructional variables into consideration. Sweller et al. (1998, p. 34) indicated that the germane load is to “separate useful, learning-relevant demands on working memory from irrelevant and wasteful forms of cognitive processing.” Therefore, germane load is relevant to the learning activities designed to improve students’ learning performance or motivation by reducing their extraneous cognitive load. Although such activities increase total cognitive load, they are helpful to learning if the total cognitive load does not exceed the students’ working memory capacity. Therefore, germane load needs to be carefully applied in designing learning activities (Kalyuga, 2009).

Consequently, learning can be effective if the cognitive loads imposed by the learning materials and by the way those materials are presented are well managed. Paas, Merriënboer, and Adam (1994) have further indicated that those three types of cognitive load can be summed up to “mental load” and “mental effort” since “extraneous cognitive load” and “germane cognitive load” are in fact two extremes that refer to the design of learning strategies. Mental load refers to the interaction among the task, subject characteristics and learning materials; it can be considered to be the load which is imposed by the task or environmental demands. Therefore, this task-centered dimension represents “intrinsic cognitive load.” Mental effort is a human-centered dimension that is concerned with the cognitive resources (i.e., the amount of resources) actually allocated to a task. Mental effort reflects the amount of learning targets and materials with which the learner is engaged, implying that it is likely to be affected by the design of the learning strategies; that is, it represents the extraneous and the germane cognitive loads (Paas, van Merriënboer, & Adam, 1994; Zheng, 2009).

In past decades, the cognitive load theory has been applied to many educational applications, including traditional instruction and technology-enhanced learning activities (Chang, Hsu, & Yu, 2011; Hollender, Hofmann, Deneke, & Schmitz, 2010; Hung et al., 2010; Liu et al., 2012). In recent years, owing to the popularity of mobile devices, mobile learning has become an important technology-enhanced learning approach; therefore, it is worth investigating the cognitive load issue in mobile learning settings. In this study, an experiment is performed by conducting an in-field activity to compare the effectiveness of mobile learning and traditional instruction approaches from the aspects of cognitive load.

Formative assessment

Formative assessment has been recognized by educators and researchers as an important element in conducting learning activities for improving student learning effectiveness (Bell & Cowie, 2001; Hargreaves, 2008; Hwang & Chang, 2011). It is a form of assessment integrated into the interaction between teachers and students for offering feedback (Oosterhof, Conrad, & Ely, 2008; Vonderwell, Liang, & Alderman, 2007). Researchers have indicated that
the provision of feedback can benefit students in improving their learning achievement and motivation (Johnson, Perry, & Shamir, 2010; Panjaburee, Hwang, Triampo, & Shih, 2010).

Owing to the advancement and popularity of computer and communication technologies, many researchers have been devoted to the development of web-based learning and assessment systems (Hwang, Tseng, & Hwang, 2008; Ho, Yin, Hwang, Shyu, & Yean, 2009). Meanwhile, formative assessments have played an important role in such web-based environments. Many researchers have indicated that the learning effectiveness of students can be improved by including the formative assessment strategy in the design of the web-learning environments (Gardner, Sheridan, & White, 2002; Gikandi, Morrow, & Davis, 2011; Wang, 2008, 2011; Wilson, Boyd, Chen, & Jamal, 2011). That is, formative assessment has been recognized as being beneficial to student learning in both traditional and web-based learning environments. Therefore, in this study, formative assessment was chosen as a representative strategy of technology-enhanced learning as opposed to conventional instructional strategies, which adopt summative assessment.

Research design

The in-field learning activity is related to the “indigenous culture” unit of the social studies course of an elementary school, and is part of the existing curriculum of the target school. The objective of the social studies course is to introduce the indigenous language, culture, and history of Taiwan. The eighteen-week course consists of in-class instruction and a field trip. Most weeks, the students receive in-class instruction from the teacher. In the middle of the semester, they visit the target ancient building for field studies. In this study, the supplementary materials and the learning tasks followed the original learning design of the course.

Participants

The participants in this experiment were two classes of fifth-grade students of an elementary school in Taiwan. One class (thirty-three students) was assigned to be the experimental group, and the other (thirty-one students) was the control group. The students were assigned to these two classes by S-shaped distribution at the beginning of the academic year. From this perspective, these two classes were viewed as having similar distributions of students in terms of their academic backgrounds.

Experiment procedure

As has been mentioned above, cognitive load theory (CLT; Paas & van Merriënboer, 1994; Sweller, van Merriënboer, & Paas, 1998) is related to the development of instructional methods to stimulate learners’ ability of applying acquired knowledge or skills to new contexts via efficiently using their limited cognitive processing capacity. Mayer and Moreno (2003) indicated that, to foster students’ ability of applying learned knowledge or skills to new situations, it is important to consider several aspects of learning design, including the presentation of learning materials, the structure of the materials, and linkage of the new knowledge or skills to the existing knowledge. In this study, a well-known e-learning strategy that has been shown to be effective by several previous studies (i.e., the formative assessment-based learning strategy) was employed for conducting mobile learning activities, in which the students were asked to observe a set of learning targets in the real-world environment based on the questions raised by the learning system. When the students failed to correctly answer some questions, the learning system did not provide the correct answers to them; instead, some clues were given to help them find the answers by themselves. We aimed to examine whether the students’ cognitive processing capacity was overloaded when they needed to face both the learning targets in the real world and the questions and clues provided by the learning systems from two aspects, mental load and mental effort. The former is related to the load imposed by the task or environmental demands, while the latter is related to the cognitive resources (i.e., the amount of resources) actually allocated to a task.

In this study, the two groups of students were situated in the same learning environment and learning content with different learning modes, that is, the conventional mobile learning mode that engaged the students in observing the learning targets and finding the answers to the learning sheet by themselves, and the traditional instructional mode in
which the teachers guided the students to observe the learning targets and presented the relevant materials of individual learning targets to them. Accordingly, we compared the loads imposed by the learning tasks and the learning achievements of the two groups.

Figure 1 shows the experimental procedure of this study. Before the field trip, an orientation was given to introduce the learning environment and the learning tasks. During the in-field learning process, the students in the experimental group were guided by a mobile device using the formative assessment-based mobile learning (FAML) strategy after receiving some training in the use of the mobile learning system. On the other hand, the students in the control group learned with the traditional in-field instruction mode; that is, they were instructed by the teacher when observing the learning targets in the field. Moreover, they received advice from the teacher when encountering problems completing the learning tasks. During the learning process, the treatments of the experimental and control groups, including the learning sheet, learning content, instructor, and learning time, were equivalent.

![Figure 1. Experiment procedure](image)

Before and after participating in the experiment, the learning achievements of the two groups of students in the social studies course were compared based on their pre-test and post-test scores. In this study, the results of the most recent mid-term examination of the social studies course were taken as the pre-test results. After participating in the experiment, the two groups of students took a post-test on the topic learned in this tour-based learning activity. Moreover, the cognitive loads of the two groups of students were analyzed and compared as well.

**Measuring tools**

The measuring tools in this study included learning achievement test sheets and questionnaires for evaluation of the learning situation. A pre-test and a post-test were developed to evaluate the learning effectiveness of the students. The pre-test was developed based on what the students had learned in the social studies course before participating in the experiment; that is, it was designed for testing the students’ prior knowledge by excluding the part of the content learned during the experiment process. It was composed of 30 multiple-choice items (60%), 10 true-or-false (20%) questions, and 3 fill-in-the-blank items (20%), with a full score of 100. On the other hand, the post-test was developed based on what they had learned in the experiment; that is, it was designed for evaluating their learning outcomes of the instructional process in this study. The post-test consisted of 40 multiple-choice items with a full score of 100. Both the pre-test and post-test were designed by the teacher who taught the indigenous culture course to the two groups of students and were evaluated by two social-science educators for expert validity.
In addition, the cognitive load survey proposed by Sweller et al. (1998) was adopted for measuring the cognitive load of individual students. The cognitive load survey consists of four items on a seven-point scale: two items for mental load (intrinsic load) and two for mental effort (extraneous and germane load). Mental load is relevant to the number of interacting information elements and the extent to which these elements interact, while mental effort is caused when learning activities and/or materials encourage higher-order thinking and challenge the learner at an appropriate level within what Vygotsky (1978) called their zone of proximal development. In this study, the Cronbach’s α value of the cognitive load form was 0.92. For the mental effort and the mental load dimensions, the Cronbach’s α values were 0.86 and 0.85, respectively. These values show high reliability.

**Learning environment**

The learning environment of this study is Chin-An temple in Tainan County of Taiwan, as shown in Figure 2. The teaching content consisted of five main parts, including the temple square, front hall, main hall, and side temples. This learning activity was designed to include twelve main units, such as the Holland wells, the stone-carved lions at the main gate and the ridge of the roof, each of which has its own architectural characteristics and historical story.

![Figure 2. Scenario of the mobile learning activity](image)

A mobile learning system using the FAML strategy was developed to conduct the local cultural learning activity with wireless networks and mobile devices, as shown in Figure 3. This system is implemented with MSSQL, ASP.NET, and IIS. There are several management functions developed for teachers, including user profile management, subject materials management, item bank management, and learning portfolio management.

![Figure 3. Framework of the Mobile Learning System](image)
Formative assessment-based mobile learning system

The FAML strategy originated from the idea of the web-based assessment and test analysis system (WATA), which is an evaluative strategy with three characteristics, including repeated answering, non-answer provision, and immediate feedback (Wang, Wang, Wang, Huang, & Chen, 2004). These characteristics allow students to repeatedly participate in the “practicing, reflecting, and revising” process. That is, students will practise and observe more, gain feedback immediately (reflect), and revise their answers (Wang et al., 2004).

From the previous applications in web-based learning environments, WATA has been recognized as being an effective formative assessment strategy. It assists learners in identifying their learning flaws to trigger their motivation for active learning, enabling them to become familiar with the learning content (Wang, 2008, 2010; Wang, Wang, & Huang, 2008). Therefore, it is interesting and appropriate to examine the effect of applying such a strategy on the learning achievements and cognitive loads of students in a mobile learning environment.

As shown in Figure 4, when students approach target positions, the mobile learning system will guide them to observe the target learning objects and interact with them based on the FAML mechanism. Once the students submit their answers, the system provides them with hints or supplementary materials, but the validity of the answers is not confirmed. That is, the students need to find out whether their answers are correct or not by themselves. For each
learning target, there are four or five test items to be answered. The students need to pass all of the items to complete the learning task of the target.

The interface of the mobile learning system is shown in Figure 5, and includes a material-browsing interface (the left snapshot) and an assessment interface (the right snapshot). After the students browse the learning materials, the learning system randomly selects test items from the item pool of that unit. Whether the answer submitted by the students is correct or not, no feedback concerning the correctness of the answer is given; instead, only hints or supplemental materials are presented. Once an item has been successively and correctly answered three times, it is removed from the item pool; however, once an incorrect answer is given, the accumulative total of correctness is reset. When the item pool of a unit is empty (i.e., all of the questions in that pool have been successively and correctly answered three times), the student is considered to have passed that unit. In this study, each student needed to pass twelve units of the indigenous culture course.

Results and analysis

Learning achievement

Before the experiment, the most recent mid-term test scores of the two groups in the social studies course were taken as the result of the pre-test, to ensure that the two groups of students had equal abilities in this subject before the learning activity. The mean and standard deviation of the pre-test for the experimental group ($M = 88.45, SD = 16.56$) and for the control group ($M = 89.48, SD = 16.01$) revealed that there was no significant difference between the two groups with $t(64) = 0.257$ and $p > .05$. That is, the two groups of students had statistically equivalent abilities before learning the subject unit.

After participating in the learning activity, the two groups of students took a post-test. The $t$-test result, $t(64) = 3.34$ and $p < .01$, showed that the learning achievement of the control group ($M = 64.60, SD = 13.37$) was better than that of the experimental group ($M = 52.88, SD = 14.61$), as shown in Table 1. The Levene’s test for equality of variances showed that the groups were homogenous with $F(1, 64) = 0.882$ and $p = .351$. Moreover, the Cohen’s $d$ value of 0.84 indicates a large effect size (Cohen, 1988). That is, a significant difference was found in the learning achievements of the two groups of students.

Such a finding (i.e., the control group had significantly better learning achievements than the experimental group) was completely different from what the researchers had expected. This surprising and interesting result indicated that the formative-assessment learning strategy for mobile learning had an unfavorable effect on the learning
achievements of the students in comparison with that for the traditional approach. Moreover, the finding is quite different from those reported by other researchers who have applied the formative assessment strategy to web-based learning (Costa, Mullan, Kothe, & Butow, 2010; Crisp & Ward, 2008; Pachler, Daly, Mor, & Mellar, 2010).

### Table 1. The t-test results for the post-test scores of the two groups of students

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
<th>Cohen’s d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control group</td>
<td>31</td>
<td>64.60</td>
<td>13.37</td>
<td>3.34***</td>
<td>0.84</td>
</tr>
<tr>
<td>Experimental group</td>
<td>33</td>
<td>52.88</td>
<td>14.61</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .01

### Cognitive load

To analyze the factors that might affect the students’ learning achievements, the cognitive loads of the two groups of students were measured by employing the cognitive load survey developed by Sweller et al. (1998) using an independent *t*-test. The result is shown in Table 2. The overall evaluation form of cognitive load and the level of mental effort (extraneous cognitive load and germane cognitive load) showed no significant difference, but the level of mental load (intrinsic load) did show a significant difference, with the degree of mental load of the experimental group being significantly higher than that of the control group.

Cognitive-load theory (Kirschner, 2002; Paas & van Merriënboer, 1994; Sweller et al., 1998) assumes that the working memory architecture (Baddeley & Logie, 1999) can be viewed as a system of limited capacity; therefore, only a limited number of elements can be handled at the same time. If the cognitive load of a task (or various tasks) exceeds the limits of working memory, then performance is seriously affected (DeStefano & LeFevre, 2007). We found no significant difference between the mental effort of the experimental group (*M* = 4.48, *SD* = 2.10) and that of the control group (*M* = 4.64, *SD* = 1.60). This implies that the instructional variables had been well considered in this formative assessment-based mobile learning strategy, which revealed a good balance between controlling the germane cognitive load and the extraneous cognitive load of the students. Therefore, the use of the formative assessment-based method (i.e., WATA) for mobile learning is an acceptable instructional design.

On the other hand, the mental load of the experimental group (*M* = 2.62, *SD* = 1.60) is significantly higher than that of the control group (*M* = 1.58, *SD* = 1.11) with *t*(64) = −3.08 and *p* < .01. That is, the average intrinsic cognitive load (task-centered dimension) for the students in the experimental group was higher than expected, which might have led to negative effects on their learning achievements. Therefore, the researchers re-examined the students’ learning portfolios and found that the experimental group students needed to answer at least twelve questions in each subject unit. Even for those students who always gave correct answers, they need to correctly answer each question three times for passing the exam. During the field trip, the students need to answer forty-eight questions and might receive three hints for guiding them to find the answer of each question via observing the real-world objects. This implies that they might be asked to observe the real-world learning objects and search for the digital-world resources at least one hundred forty-four times. Because it takes much more time to seek the relevant learning materials in a mobile learning environment that combines both real-world and digital-world resources than in a pure web-based learning environment, the excessive amount of learning content (questions or test items for guiding the students to observe and learn) could be the main reason for the excessive mental load and this unexpected result.

### Table 2. The *t*-test results for the cognitive load of the two groups of students

<table>
<thead>
<tr>
<th>Cognitive Load</th>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mental effort</td>
<td>Control</td>
<td>31</td>
<td>4.64</td>
<td>2.10</td>
<td>0.33</td>
</tr>
<tr>
<td>(Extraneous load and germane load)</td>
<td>Experimental</td>
<td>33</td>
<td>4.48</td>
<td>1.60</td>
<td></td>
</tr>
<tr>
<td>Mental load</td>
<td>Control</td>
<td>31</td>
<td>1.58</td>
<td>1.11</td>
<td>−3.08**</td>
</tr>
<tr>
<td>(Intrinsic load)</td>
<td>Experimental</td>
<td>33</td>
<td>2.62</td>
<td>1.60</td>
<td></td>
</tr>
</tbody>
</table>

** *p* < .01
Conclusions and discussion

This study explores the effect of online learning strategies on a mobile learning environment that combines digital learning resources and real-world learning contexts. A mobile learning activity was conducted with a formative assessment-based learning strategy; moreover, we investigated the cognitive loads and learning achievements of the students to examine the effect of the learning strategy on such an innovative and complex learning scenario.

The results showed that the students in the experimental group did not achieve as much as those in the traditional group. This finding is quite different from what has been previously reported (Chu, Hwang, & Tsai, 2010; Chen et al., 2010; Hwang & Chang, 2011; Hwang et al., 2009; Wang, 2008). Further analyses revealed that one of the reasons why students who learned in the real-world scenario with mobile devices obtained unfavorable learning achievements was their high cognitive loads. From the t-test result of the cognitive load, it was found that high mental load was the main factor causing the negative effect on the learning achievements in the experimental group. This implies that the use of the formative assessment strategy was not the main reason for this unexpected result, since mental load refers to the overloading of the working memory of students. Because the students were repeatedly asked to find learning materials from both the real-world and the digital world environments in a limited period of time and there were many learning tasks to be completed, students’ working memories were likely to be overloaded.

From the observations of the students’ in-field learning behaviors, it was found that most were eager to complete the learning tasks. Because time was limited, students tended to answer the questions in a hurry and therefore were likely to give incorrect answers. When they did, they were asked to make additional observations and answer the questions again. On the other hand, the feedback from the learning system was instant and frequent, which could make the students become anxious since they needed to read the hints or guidance repeatedly and, most of the time, the feedback might relate to an additional observation to be made in the present learning task. Adding to their pressure, some of their peers might have already proceeded to the next learning task. Therefore, most of the students hurried to repeatedly observe the learning targets and answer the questions. In that case, they did not make deep reflections or thorough observations before going on to the next learning task, which not only caused repeated failures in the tasks, but also led to high cognitive loads.

It is also interesting to compare the result of this study with those of previous studies (e.g., Costa et al., 2010; Crisp & Ward, 2008; Pachler et al., 2010; Wang, 2008, 2010; Wang, Wang, & Huang, 2008) that reported the effectiveness of the formative assessment-based learning strategy in web-based learning environments. From this case, we can understand that using mobile devices to learn in an authentic learning environment is not always successful, even when adopting those learning strategies that have been reported as being effective in the well-recognized web-based learning environments. Therefore, it is important to investigate whether the learning strategies that have been claimed to be effective in traditional computer systems or web-based learning environments can work well for mobile or ubiquitous learning. Moreover, it is also important to develop new strategies or modify existing strategies for learning design by taking the special features of mobile and ubiquitous learning (i.e., interacting with real-world contexts with personalized supports or hints from the digital system) into account. This conclusion conforms to what has been reported by Chu, Hwang, Tsai, and Tseng (2010), who claim that developing proper learning guidance procedures or tools will help students improve their learning achievements in a mobile and ubiquitous learning environment. That is, with a proper learning design, the effect of mobile and ubiquitous learning can be much better than that of the traditional approach.

These findings also show that decreasing students’ cognitive load could be another important issue for mobile learning. Most web-based learning studies have devoted their efforts to building virtual reality tools or learning materials to help the students construct knowledge (Verhoeven, Schnotz, & Paas, 2009; Wang et al., 2004; Wang et al., 2008). Nevertheless, in mobile learning environments, the students are already situated in a more complex learning situation that combines real-world and digital-world resources; therefore, it is a challenging issue to propose new learning strategies or rethink the existing approaches such that the students can learn with reasonable cognitive loads.

To sum up, the findings of this study are helpful to those researchers who plan to develop mobile learning systems or conduct mobile learning activities in the future. For those researchers who are interested in this study, the authors would be happy to provide the experimental data upon request. In the meantime, since this study focuses mainly on a fifth-grade social studies course in an elementary school, it is worth investigating the effects of similar approaches on
other subjects and different grades of students. Currently, we are working on extending this study by conducting various mobile learning activities for some natural science courses in several elementary schools.

Acknowledgements

The author would like to thank Ms. Yen-Ru Shih and Ms. Hsun-Fang Chang for their assistance in developing the learning materials and conducting the experiment. This study was supported in part by the National Science Council of the Republic of China under contract number NSC 102-2511-S-031-001 and NSC 101-2628-S-031-001-MY2.

References


Simulation and Learning: A Model-Centered Approach
(Book Review)

Reviewer:
Dirk Ifenthaler
Professor, Deakin University, Australia
Director, Centre for Research in Digital Learning
dirk@ifenthaler.info

Textbook Details:
Simulation and Learning: A Model-Centered Approach
Written by Franco Landriscina

Simulation and Learning: A Model-Centered Approach’ provides a comprehensive and state-of-the-art review on the nature of simulation from cognitive and technological perspectives as well as suggests practical implications for effective integration of simulation into learning and instruction. The theory of mental models functions as the theoretical foundation, using a wide range of mental model research such as Stachowiak’s neo-pragmatic and Johnson-Liard’s cognitive-constructivist approaches as well as Seel’s model-centered learning concept. Key topics covered by the author include foundations of simulation, simulation-based learning, dynamic systems modeling, and simulation-based instruction.

As more and more simulations are being implemented for very specific purposes, classification systems for simulation are scarce and different theoretical concepts provide diverse explanations of cognitive processes regarding simulation. Still, educational training programs using simulation have been successfully applied in the fields of flight training, health care education, dental education, command and control training of large incidents, team-based decision making, simulations for the training of firefighters, teacher training, and many other domains. However, comprehensive frameworks linking cognitive processes, learning, instruction, and simulation are scarce.

The significance of this monograph is the emphasis on the learners’ cognitive functions and their link to the computer-based application while learning with simulation. This perspective is rare within the well-established research on simulation in educational contexts and adds an important conceptual view to mental model theory. Besides computer-based modelling and simulation tools for the construction of models, there are also tools which focus on the exploration of the underlying model. This second type of computer-based modelling tools for the exploration of underlying models may be differentiated further into (1) black-box and (2) glass-box systems. In a black-box simulation all computations are hidden from the learner. Such systems include adjustable parameters and a large variety of process visualizations which become available after the actual action of the simulation is completed. On the other hand, glass-box simulations expose the underlying mathematical and logical simulation model. This can be realized by revealing the underlying equations, connections, and interdependencies between variables and changing variables, parameters, or equations. Based on this important conceptual view of simulation and cognitive processes, the author offers implications for the design, development, and implementation of simulation in formal learning contexts.

The book is divided into seven chapters. Chapter 1, entitled ‘An Introduction to Simulation for Learning’, introduces the reader to simulation for learning. The chapter offers a critical review of the epistemic status of simulation, provides multiple definitions of simulation, and examines the relationship between simulation and learning. Finally, the chapter suggests the need for a multidisciplinary approach when investigating the potentials of simulation for learning. Chapter 2, entitled ‘Simulation and Cognition’, presents an overview of cognitive theories which use simulation as a metaphor for cognitive processing. The theory of mental models is used for explaining the complex functions of information processing and its internal and external representations. Chapter 3, entitled ‘Models Everywhere’, investigates the epistemic role of models and the various types of representation of a (cognitive) system. A taxonomy of models suggests numerous types of models, such as, physical, visual, linguistic, and electrical models. Chapter 4, entitled ‘Simulation Modeling’, claims that the fundamental requirement of a simulation is a model, of a real or imagined system. Detailed examples regarding the modeling and simulation process as well as applications from dynamical systems modeling, continuum physics modeling, molecular
dynamics, compartmental models, agent-based modeling, system dynamics, as well as cellular modeling and simulation are explained and contrasted. Chapter 5, entitled ‘Simulation-Based Learning’, provides examples how people learn with simulations. Again, the underlying cognitive processes of simulation-based learning are illustrated using examples from various disciplines. Additionally, a comprehensive view on model-based learning and teaching as well as links to cognitive load theory and instructional design is provided in the concluding sections. Chapter 6, entitled ‘Simulations for Thinking’, examines the relation between mental simulation and computer-based simulation. Similarities and differences of these types of simulation are discussed. The argument is built mainly around cognitive partnering, analogical thinking, as well as simulation and language. Chapter 7, entitled ‘Simulation-Based Instruction’, highlights how simulation can be applied for teaching. Examples are drawn from cross-curricular learning environments and provide the reader a rich collection of practical examples for teaching with simulation. The book concludes with the Epistemic Cycle, a model which describes how learners construct knowledge through simulation in a formal learning context.

The Appendix includes a well-established collection of ‘Simulation Resources’ for equation-based modeling, molecular dynamics, epidemiological modeling, agent-based modeling, system dynamics, cellular modeling and simulation, as well as systems modeling and design. The collection of over 35 resources includes the name of the application as well as a corresponding link.

One of the main strengths of this book are the numerous examples of well-designed learning environments using simulation. However, authentic tasks for simulation are required that enable the learners to explore the environment in dynamic interaction with the context as if they were really present. Furthermore, a stronger emphasis on embedded supports and scaffolds could be provided to the reader, as these pedagogical interventions assist novice and expert problem solvers in operating within the simulation learning environment.

Overall, ‘Simulation and Learning: A Model-Centered Approach’ provides a rare theoretical perspective on simulation and cognitive processes. Additionally, it offers many good models from educational practice in various subject domains. The book is easy to follow as the many examples provide practical insights into instructional design with simulation. In this respect, it caters different expectations of educational technologists, instructional designers, and course facilitators with regard to finding theoretical understanding of simulation as well as implementing simulation into formal learning context.